

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:17:33 ; Search time 26,1818 Seconds
(without alignments)
45,805 Million cell updates/sec

Title: US-09-646-532B-3

Perfect score: 46

Sequence: 1 NFRYTNFAX 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.101002:*

- 1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 45 | 97.8 | 9 | 20 | AAV42755 |
| 2 | 45 | 97.8 | 9 | 20 | AAV39333 |
| 3 | 38 | 82.6 | 28 | 12 | AA14366 |
| 4 | 34 | 73.9 | 316 | 22 | ABG24752 |
| 5 | 33 | 71.7 | 436 | 22 | AAU4256 |
| 6 | 33 | 71.7 | 1180 | 22 | ABG22442 |
| 7 | 32 | 69.6 | 76 | 22 | AAU33056 |
| 8 | 32 | 69.6 | 209 | 22 | AAU33056 |
| 9 | 32 | 69.6 | 245 | 21 | AAU33056 |
| 10 | 32 | 69.6 | 288 | 21 | AAU33056 |

| | | | | | |
|----|----|------|------|----|----------|
| 11 | 32 | 69.6 | 327 | 21 | AAU33056 |
| 12 | 32 | 69.6 | 372 | 21 | AAU33056 |
| 13 | 32 | 69.6 | 392 | 22 | AAU33056 |
| 14 | 32 | 69.6 | 402 | 23 | AAU33056 |
| 15 | 32 | 69.6 | 511 | 21 | AAU33056 |
| 16 | 32 | 69.6 | 550 | 21 | AAU33056 |
| 17 | 32 | 69.6 | 595 | 21 | AAU33056 |
| 18 | 32 | 69.6 | 621 | 21 | AAU33056 |
| 19 | 32 | 69.6 | 850 | 17 | AAU33056 |
| 20 | 32 | 69.6 | 859 | 15 | AAU33056 |
| 21 | 32 | 69.6 | 927 | 23 | AAU33056 |
| 22 | 32 | 69.6 | 1147 | 14 | AAU33056 |
| 23 | 32 | 69.6 | 1178 | 18 | AAU33056 |
| 24 | 32 | 69.6 | 1181 | 15 | AAU33056 |
| 25 | 32 | 69.6 | 1181 | 15 | AAU33056 |
| 26 | 32 | 69.6 | 1181 | 16 | AAU33056 |
| 27 | 32 | 69.6 | 1181 | 17 | AAU33056 |
| 28 | 32 | 69.6 | 1183 | 18 | AAU33056 |
| 29 | 32 | 69.6 | 1338 | 19 | AAU33056 |
| 30 | 32 | 69.6 | 1666 | 22 | AAU33056 |
| 31 | 31 | 67.4 | 9 | 16 | AAU33056 |
| 32 | 31 | 67.4 | 39 | 22 | AAU33056 |
| 33 | 31 | 67.4 | 39 | 22 | AAU33056 |
| 34 | 31 | 67.4 | 39 | 22 | AAU33056 |
| 35 | 31 | 67.4 | 39 | 22 | AAU33056 |
| 36 | 31 | 67.4 | 187 | 22 | AAU33056 |
| 37 | 31 | 67.4 | 187 | 23 | AAU33056 |
| 38 | 31 | 67.4 | 191 | 23 | AAU33056 |
| 39 | 31 | 67.4 | 376 | 17 | AAU33056 |
| 40 | 31 | 67.4 | 377 | 16 | AAU33056 |
| 41 | 31 | 67.4 | 377 | 16 | AAU33056 |
| 42 | 31 | 67.4 | 377 | 17 | AAU33056 |
| 43 | 31 | 67.4 | 377 | 21 | AAU33056 |
| 44 | 31 | 67.4 | 377 | 23 | AAU33056 |
| 45 | 31 | 67.4 | 377 | 23 | AAU33056 |

ALIGNMENTS

| | | |
|----------|--|-------------------------------|
| RESULT 1 | AAV42755 | standard: peptide; 9 AA. |
| ID | AAV42755 | |
| AC | AAV42755 | |
| DF | 20-DEC-1999 | (first entry) |
| DE | Wheat amyloplast ADP-glucose transporter peptide #3. | |
| XX | Starch biosynthesis; amyloplast; ADP-glucose; transport; import; | |
| KW | amylopectin; amylose; branching; chemical structure; transgenic plant; | |
| OS | optimisation; industrial applications. | |
| XX | Triticum aestivum. | |
| XX | Key | Location/Qualifiers |
| FT | Misc-difference 9 | /label= Xaa |
| FT | | /note= "Xaa = any amino acid" |
| XX | WO9947682-A1. | |
| XX | 23-SEP-1999. | |
| PD | 19-MAR-1999; | 99WO-GB00728. |
| PF | 20-MAR-1998; | 98GB-0005939. |
| PR | (UYMA-) UNIV VICTORIA MANCHESTER. | |
| XX | Emes MJ, Tellow IJ, Bowsheer CG; | |

Arabidopsis thalia
Arabidopsis thalia
Novel human secret
Herbicideally activ
Arabidopsis thalia
Arabidopsis thalia
Recombinant UB-A54
Helicobacter pylori
120-128 K Dalton
H. pylori taga ant
Chimeric receptor
CAT antigen. Heli
H. pylori cytoplas
Tag A antigen of H
H. pylori taga ant
Helicobacter pylori
H. pylori cytoplas
Helicobacter pylori
Drosophila melanog
Bovine GGTase-I be
Peptide #10381 enc
Human brain expres
Human bone marrow
Peptide #10729 enc
Human G protein-co
Novel G protein co
Human FSH-like GPC
Rat geranylgeranyl
Rat GGTase-I beta-
Human GGTase-I bet
Human geranylgeran
Human geranylgeran
Pain regulated pro
Pain regulated pro

DR WPI: 1999-590977/50.
 XX
 PT New transporter complex protein useful for modulating starch content in
 PT plants, especially useful in food production -
 XX
 PS Claim 1; Page 3; 28pp; English.
 XX
 CC This sequence represents a wheat amyloplast ADP-glucose transporter
 CC peptide, #2. The wheat amyloplast ADP-glucose transporter is associated
 CC with the amyloplast membrane and comprises at least two proteins; this
 CC sequence, along with peptides #1 (AAV42753) and #2 (AAV42754) are
 CC components of one of these proteins. Peptides #4-#7 (AAV42756-42759)
 CC represent sequences within the second protein (AAV42760). The sugar
 CC nucleotide ADP-glucose is the immediate substrate for starch synthesis,
 CC which occurs in the amyloplast; however, ADP-glucose is mainly
 CC synthesised outside the amyloplast in the cytoplasm. The ADP-glucose
 CC transporter is responsible for the import of ADP-glucose into the
 CC amyloplast and therefore plays a pivotal role in the regulation of starch
 CC synthesis. The transporter not only influences starch yield, but also
 CC quality as the starch synthases involved in amylose and amylopectin
 CC synthesis have different affinities for ADP-glucose. Variations in the
 CC chemical structure of starch are determined by the ratio of amylose to
 CC amylopectin, and by the degree of branching in amylopectin in the starch
 CC polymer. These variations can significantly alter the properties of
 CC starch. The ADP-glucose transporter complex is useful for generating
 CC transgenic plants in which the starch quality can be optimised for
 CC industrial applications in which starch is used. For example, transgenic
 CC plants which have an increased amylose content in starch are useful for
 CC production of starch with increased viscosity and gel strength, which
 CC prevents baked food going stale so quickly. Conversely, an increased
 CC amylopectin content in the starch produces waxy starch useful as
 CC thickening agents in food and coatings.
 XX
 SQ Sequence 9 AA:
 Query Match 97.8%; Score 45; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFRYTNFA 8
 |||||||
 Db 1 NFRYTNFA 8
 RESULT 2
 AAV39333
 ID AAV39333 standard; peptide: 9 AA.
 XX
 AC AAV39333;
 XX
 DT 01-DEC-1999 (first entry)
 XX
 DE ADP glucose transporter peptide 3.
 XX
 KM ADP glucose transporter; transform plant cell; wheat; starch production;
 KM waxy starch; thickening agent; food; coating; increased viscosity; stale;
 KM gel strength; baked food.
 XX
 OS Trilicium aestivum.
 OS
 FH Key Location/Qualifiers
 FT Misc-difference 9 /note= "Any amino acid"
 FT
 XX W09947681-A1.
 XX
 PD 23-SEP-1999.
 XX
 PF 19-MAR-1999; 99WO-GB00727.
 XX
 PR 20-MAR-1998; 98GB-0005939.
 XX
 PA (UYMA-) UNIV VICTORIA MANCHESTER.

XX
 PI Emes MJ, Tetlow IJ, Bowsher CG;
 XX
 DR WPI: 1999-571841/48.
 XX
 PT ADP glucose transporter protein used for modifying plant starch
 PT production -
 XX
 PS Claim 1; Page 15; 26pp; English.
 XX
 CC Peptides AAV39331-39337 are fragments of an ADP glucose transporter
 CC protein. The protein contains at least one of the peptide sequences, and
 CC is capable of ADP glucose transport. A DNA molecule encoding an ADP
 CC glucose transporter protein can be used to transform plant cells. The
 CC ADP glucose transporter protein can be used to regulate starch
 CC production from a plant. The plants can then be used to produce waxy
 CC starches that can be used as thickening agents in food and coatings.
 CC Alternatively the plants can be used to produce a starch with increased
 CC viscosity and gel strength, which can be incorporated in baked food which
 CC stays fresh for a longer length of time.
 XX
 SQ Sequence 9 AA:
 Query Match 97.8%; Score 45; DB 20; Length 9;
 Best Local Similarity 100.0%; Pred. No. 7.8e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFRYTNFA 8
 |||||||
 Db 1 NFRYTNFA 8
 RESULT 3
 AAR14366
 ID AAR14366 standard; Peptide: 28 AA.
 XX
 AC AAR14366;
 XX
 DT 30-JAN-1992 (first entry)
 XX
 DE N-terminal fragment of C7MA 28A32 36K protein.
 XX
 KM Tumour cell; epitope; colon; cancer; monoclonal antibodies; vaccine;
 KM Antigen; carcinoma.
 XX
 OS Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT Region 1 /label= unknown
 FT Region 26 /label= unknown
 FT Region /label= unknown
 XX
 FN W09116629-A.
 XX
 PD 31-OCT-1991.
 XX
 PF 10-APR-1991; 91WO-U002459.
 XX
 PR 12-APR-1990; 90US-0508373.
 XX
 PA (ALKU) AKZO NV.
 XX
 PI Pomato N, Bos ES, Ransom JH, Hanna MG;
 XX
 DR WPI: 1991-339972/46.
 XX
 PT New human Mab 28A32 - used to diagnose and monitor colon cancer
 PT and to prepare vaccines.
 XX
 PS Claim 1; Fig 7; 34pp; English.
 XX
 CC The sequence was obtd. by Edman degradation and represents the

CC most predominant and reproducible isolatable fragment of the 36K
CC protein. The protein is one of four (50K, 46K, 36K, and 32K)
CC associated with the CTA 28A32 antigen. The 36K protein is
CC associated with membranes; the rest are found in the cytoplasm of
CC cells. The antigen is associated with colon tumours and mono-
CC clonal antibody 28A32 can be used to monitor and detect the
CC presence of tumour cells. Vaccines can also be prepd. from the
CC proteins.
CC See also AAR14364-67.
XX
SQ Sequence 28 AA;
Query Match 82.6%; Score 38; DB 12; Length 28;
Best Local Similarity 87.5%; Pred. No. 2.4;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 NFRYTNFA 8
DB 15 NFDYTNFA 22
II IIII
II IIII
RESULT 4
ABG24752
ID ABG24752 standard; Protein; 316 AA.
XX
AC ABG24752;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #24743.
XX
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HISE-) HISEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS88939.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 5111; 103bp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 316 AA;
Query Match 73.9%; Score 34; DB 22; Length 316;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 NFRYTNFA 8
DB 195 NFRYKDEFA 202
II IIII
II IIII
RESULT 5
AAU42256
ID AAU42256 standard; Protein; 436 AA.
XX
AC AAU42256;
XX
DT 27-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein #3152.
XX
KW SAPHO syndrome; synovitis; acne; pustulosis; hyperostosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX
OS Propionibacterium acnes.
XX
PN WO200181581-A2.
XX
PD 01-NOV-2001.
XX
PF 20-APR-2001; 2001WO-US12865.
XX
PR 21-APR-2000; 2000US-199047P.
PR 02-JUN-2000; 2000US-208841P.
PR 07-JUL-2000; 2000US-216747P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D.
XX
DR WPI: 2001-616774/71.
DR N-PSDB: AAS59516.
XX
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris -
XX
PS Example 1; SEQ ID No 3451; 1069bp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to

CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

SO Sequence 436 AA;

Query Match 71.7%; Score 33; DB 22; Length 436;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 FRYTNEA 8
||| |
Db 369 FRYTGFA 375

RESULT 6
ABG22442
ID ABG22442 standard; Protein: 1180 AA.

AC ABG22442;
DT 18-FEB-2002 (first entry)
DE Novel human diagnostic protein #22433.

KW Human: Chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.
PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI: 2001-639362/73.
DR N-PSDB: MAS86629.

PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

PS Claim 20; SEQ ID No 52801; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pcl_sequences.

SO Sequence 1180 AA;

Query Match 71.7%; Score 33; DB 22; Length 1180;
Best Local Similarity 71.4%; Pred. No. 7.8e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
||| |
Db 47 NFRYNNY 53

RESULT 7
AAU33056
ID AAU33056 standard; Protein: 76 AA.

AC AAU33056;
DT 18-DEC-2001 (first entry)
DE Novel human secreted protein #3547.

KW Human: vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

OS Homo sapiens.
PN WO200179449-A2.

PD 25-OCT-2001.

PF 16-APR-2001; 2001WO-US08656.

PR 18-APR-2000; 2000US-0552929.
PR 26-JAN-2001; 2001US-0770160.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-611725/70.

PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy -
PS Claim 20; Page 702; 765pp; English.

XX The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising
CC the nucleic acids encoding the polypeptides and cells genetically
CC engineered to express them are also useful for producing the proteins.
CC The proteins are useful in genetic vaccination, testing and
CC therapy, and can be used as nutritional supplements. They may be used to
CC increase stem cell proliferation; to regulate haematopoiesis; and in
CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
CC immune suppression and/or stimulation; as anti-inflammatory agents; and
CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
CC sequences of novel human secreted proteins of the invention.

SO Sequence 76 AA;

Query Match 69.6%; Score 32; DB 22; Length 76;
Best Local Similarity 83.3%; Pred. No. 82;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 FRYTNF 7
1:||||
DB 57 FRYTNF 62

RESULT 8
AAG84920
ID AAG84920 standard; Protein: 209 AA.

AC AAG84920;

DT 11-SEP-2001 (first entry)

DE Shrimp white spot Bacilliform virus (WSBV) protein 11.

KW Shrimp white spot Bacilliform virus; MSBV; diagnosis; viral infection;
antiviral agent; gene expression; antisense construct;

KM transgenic viral resistant shrimp.

OS White spot syndrome virus.

PN MO200138351-A2.

PD 31-MAY-2001.

PF 08-NOV-2000; 2000WO-US28888.

PR 24-NOV-1999; 99CN-0124717.

PA (PENY-) PE CORP NY.
(THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
(SINO-) SINOGENOMAX CO LTD.

PI Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;

DR WPI: 2001-355877/37.

PS N-PSDB: AAG62700.

PT Primary nucleotide sequence of the shrimp white spot Bacilliform virus
(MSBV), useful for producing viral polypeptides that can be used to
screen for agents that are useful for treating MSBV infection -

PS Claim 1: Figure 3; 626pp; English.

CC The invention provides the primary nucleotide sequence of the WSBV genome
(AAG62689), predicted transcript sequences (AAG62689-AAG62839) and
CC encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
(AAG62840-63160) suitable for use as primers or probes. The nucleic acid
CC molecules and proteins of the invention are useful for diagnosis and
CC monitoring viral infection, in screens for antiviral agents and for
CC monitoring viral gene expression or activity during a treatment regimen.
CC The nucleic acid molecules are also useful as antisense constructs to
CC control viral gene expression in infected cells and tissues and to create
CC transgenic viral resistant shrimp.

SQ Sequence 209 AA;

Query Match 69.6%; Score 32; DB 22; Length 209;
Best Local Similarity 71.4%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRTYTNF 7
1:||||
DB 32 NFRTYTNV 38

RESULT 9
AAB07810
ID AAB07810 standard; Protein: 245 AA.

XX AAB07810;

DT 14-NOV-2000 (first entry)

DE A galactanase of Bacillus agaradhaerens AC13.

KW Galactanase; animal feed; galactan; glycosyl hydrolase; textile;
detergent; wine; juice; cellulose processing.

OS Bacillus agaradhaerens.

PN MO200047711-A2.

PD 17-AUG-2000.

PF 08-FEB-2000; 2000WO-DK00052.

PR 11-FEB-1999; 99DK-0000184.
07-JUN-1999; 99DK-0000799.

PA (NOVO) NOVO NORDISK AS.

PI Bioernvad ME, Clausen IG, Schuelein M, Bech L, Oestergaard PR;
Sjoeholm C;

DR WPI: 2000-565292/52.

PS N-PSDB: AAG59394.

PT Novel methods for modifying animal feed using galactanase and novel
galactanase enzymes useful for modifying animal feed -

PS Claim 38; Page 64-65; 77pp; English.

CC The present sequence represents a Bacillus galactanase enzymes. The
CC enzyme comprises at least one of the consensus sequences AAB07802-07,
CC and is used in the method of the invention. The specification
CC describes a method for modifying animal feed using galactanase.
CC Galactanase enzymes degrade galactans. The galactanases of
CC the invention are glycosyl hydrolases. The galactanase enzymes are
CC useful for the modification of animal feed and in the textile,
CC detergent, wine and juice and cellulose processing industries.

SQ Sequence 245 AA;

Query Match 69.6%; Score 32; DB 21; Length 245;
Best Local Similarity 71.4%; Pred. No. 2.6e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRTYTNF 7
1:||||
DB 47 NFRTYTNF 53

RESULT 10
AAG57319
ID AAG57319 standard; Protein: 288 AA.

AC AAG57319;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 73849.

KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

OS Arabidopsis thaliana.

PN EP1033405-A2.

PD 06-SEP-2000.

[illegible]

PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.

PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 69.6%; Score 32; DB 21; Length 327;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRITNF 7
||| |||
Db 115 NFRITNF 121

RESULT 12

AA657317
ID AA657317 standard; Protein; 372 AA.

XX AC AA657317;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 73847.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 16-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136382.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139482.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142380.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145228.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149724.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 27-AUG-1999; 99US-0151303.
PR 30-AUG-1999; 99US-0151438.
PR 31-AUG-1999; 99US-0151930.
PR 01-SEP-1999; 99US-0152369.
PR 07-SEP-1999; 99US-0152369.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.

PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 26-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.6%; Score 32; DB 21; Length 372;
 Best Local Similarity 85.7%; Pred. No. 3.9e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
 |||||
 Db 160 NFRYTNF 166

RESULT 13
 AAU30499
 ID AAU30499 standard; Protein; 392 AA.

AC AAU30499;

DT 18-DEC-2001 (first entry)

XX Novel human secreted protein #990.

DE Human; vaccination; gene therapy; nutritional supplement;

KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;

KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.

XX Homo sapiens.

OS WO200179449-A2.

XX 25-OCT-2001.

PD 16-APR-2001: 2001WO-US08656.

PR 18-APR-2000: 2000US-0552929.

PR 26-JAN-2001: 2001US-0770160.

XX (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI; 2001-611725/70.

XX Nucleic acids encoding a range of human polypeptides, useful in genetic

PT vaccination, testing and therapy -

XX Claim 20; Page 301-302; 765pp; English.

CC The invention relates to novel human secreted polypeptides. The

CC polypeptides and antibodies to the polypeptides are useful for

CC determining the presence of or predisposition to a disease associated

CC identifying agents (agonists and antagonists) that bind to them. Cells

CC expressing the proteins are useful for identifying a therapeutic agent

CC for use in treatment of a pathology related to aberrant expression or

CC physiological interactions of the polypeptide. Vectors comprising

CC the nucleic acids encoding the polypeptides and cells genetically

CC engineered to express them are also useful for producing the proteins.

CC The proteins are useful in genetic vaccination, testing and

CC therapy, and can be used as nutritional supplements. They may be used to

CC increase stem cell proliferation; to regulate haematopoiesis; and in

CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;

CC immune suppression and/or stimulation; as anti-inflammatory agents; and

CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid

CC sequences of novel human secreted proteins of the invention.

XX SQ Sequence 392 AA;

OY 2 FRYTNF 7
 |||||

Db 153 FRYTNF 158

RESULT 14
 ABB91033
 ID ABB91033 standard; Protein; 402 AA.

AC ABB91033;

DT 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 244.

DE Herbicidally active polypeptide SEQ ID NO 244.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

OS WO200210210-A2.

XX 07-FEB-2002.

PD 28-AUG-2001: 2001WO-EP09892.

PR 28-AUG-2001: 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M;

PI WPI; 2002-269010/31.

DR Identifying plant target proteins for herbicidally active compounds,

PT comprising aligning and comparing nucleic acid or amino acid sequences

PT from plant with nucleic acid or amino acid sequences from non-plant

PT organisms -

XX Claim 5; SEQ ID NO 244; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins

CC (ABB90790-ABB94016) for herbicidally active compounds, comprising

CC aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using

CC suitable search parameters, where plant sequences having an E-value

CC greater by a factor of 3 than the E-value of most similar non-plant

CC sequences are selected. The polypeptides or nucleic acids encoding them

CC are useful for identifying modulators. The identified modulators are

CC useful as herbicides.

XX SQ Sequence 402 AA;

Query Match 69.6%; Score 32; DB 23; Length 402;

Best Local Similarity 71.4%; Pred. No. 4.2e+02;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NRRYTNF 7
11111
Db 114 NRRYTNF 120

RESULT 15

AG58809
ID AG58809 standard; Protein: 511 AA.

XX
AC AG58809;

XX
DT 18-OCT-2000 (first entry)

XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 75981.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.

PR
PR 09-MAR-1999; 99US-0123180.

PR
PR 23-MAR-1999; 99US-0123548.

PR
PR 25-MAR-1999; 99US-0125788.

PR
PR 29-MAR-1999; 99US-0126785.

PR
PR 01-APR-1999; 99US-0127462.

PR
PR 06-APR-1999; 99US-0128234.

PR
PR 08-APR-1999; 99US-0128714.

PR
PR 16-APR-1999; 99US-0129845.

PR
PR 19-APR-1999; 99US-0130077.

PR
PR 21-APR-1999; 99US-0130449.

PR
PR 23-APR-1999; 99US-0130510.

PR
PR 28-APR-1999; 99US-0130891.

PR
PR 30-APR-1999; 99US-0131449.

PR
PR 04-MAY-1999; 99US-0132048.

PR
PR 05-MAY-1999; 99US-0132407.

PR
PR 06-MAY-1999; 99US-0132485.

PR
PR 07-MAY-1999; 99US-0132486.

PR
PR 11-MAY-1999; 99US-0132863.

PR
PR 14-MAY-1999; 99US-0134256.

PR
PR 14-MAY-1999; 99US-0134218.

PR
PR 14-MAY-1999; 99US-0134221.

PR
PR 18-MAY-1999; 99US-0134370.

PR
PR 19-MAY-1999; 99US-0134768.

PR
PR 20-MAY-1999; 99US-0134941.

PR
PR 21-MAY-1999; 99US-0135124.

PR
PR 24-MAY-1999; 99US-0135353.

PR
PR 25-MAY-1999; 99US-0135629.

PR
PR 27-MAY-1999; 99US-0136021.

PR
PR 28-MAY-1999; 99US-0136392.

PR
PR 01-JUN-1999; 99US-0136782.

PR
PR 03-JUN-1999; 99US-0137222.

PR
PR 04-JUN-1999; 99US-0137528.

PR
PR 07-JUN-1999; 99US-0137502.

PR
PR 08-JUN-1999; 99US-0137724.

PR
PR 10-JUN-1999; 99US-0138094.

PR
PR 10-JUN-1999; 99US-0138540.

PR
PR 14-JUN-1999; 99US-0138847.

PR
PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139750.

PR 21-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

Search completed: March 25, 2003, 08:19:42
 Job time : 27.1818 secs

PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158359.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 69.68; Score 32; DB 21; Length 511;
 Best Local Similarity 85.7%; Pred. No. 5.3e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
 ||| |||
 Db 76 NFRYTNF 82

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:22:27 ; Search time 8.45455 Seconds

(without alignments)
56.911 Million cell updates/sec

Title: US-09-646-532b-3

Perfect score: 46

Sequence: 1 NFRYTNFA 9

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published_Applications_AA:*

1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length DB | ID | Description |
|------------|-------|-------------|-----------|----|---------------------|
| 1 | 32 | 69.6 | 40 | 12 | US-10-001-843-127 |
| 2 | 32 | 69.6 | 927 | 12 | US-10-003-356-8 |
| 3 | 32 | 69.6 | 1338 | 10 | US-09-402-100-4 |
| 4 | 31 | 67.4 | 39 | 10 | US-09-864-761-45646 |
| 5 | 31 | 67.4 | 187 | 10 | US-09-811-284-228 |
| 6 | 31 | 67.4 | 311 | 10 | US-09-928-175-15 |
| 7 | 31 | 67.4 | 359 | 10 | US-09-928-175-10 |
| 8 | 31 | 67.4 | 383 | 10 | US-09-928-175-5 |
| 9 | 31 | 67.4 | 646 | 10 | US-09-928-175-13 |
| 10 | 31 | 67.4 | 682 | 10 | US-09-928-175-12 |
| 11 | 31 | 67.4 | 694 | 10 | US-09-928-175-8 |
| 12 | 31 | 67.4 | 713 | 9 | US-09-965-536A-6 |
| 13 | 31 | 67.4 | 718 | 10 | US-09-928-175-3 |
| 14 | 31 | 67.4 | 730 | 10 | US-09-928-175-7 |
| 15 | 31 | 67.4 | 737 | 9 | US-09-965-536A-2 |
| 16 | 31 | 67.4 | 754 | 10 | US-09-928-175-2 |
| 17 | 31 | 67.4 | 1163 | 10 | US-09-826-660-21 |
| 18 | 31 | 67.4 | 1193 | 10 | US-09-873-873-30 |
| 19 | 30 | 65.2 | 456 | 10 | US-09-925-300-1595 |

| | | | | | | |
|----|----|------|-----|----|--------------------|---------------------|
| 20 | 30 | 65.2 | 711 | 10 | US-09-862-658-2 | Sequence 2, Appl 1 |
| 21 | 29 | 63.0 | 198 | 9 | US-10-104-019-38 | Sequence 38, Appl 1 |
| 22 | 29 | 63.0 | 199 | 9 | US-10-104-019-32 | Sequence 34, Appl 1 |
| 23 | 29 | 63.0 | 199 | 9 | US-10-104-019-33 | Sequence 33, Appl 1 |
| 24 | 29 | 63.0 | 199 | 9 | US-10-104-019-34 | Sequence 35, Appl 1 |
| 25 | 29 | 63.0 | 199 | 9 | US-10-104-019-35 | Sequence 36, Appl 1 |
| 26 | 29 | 63.0 | 199 | 9 | US-10-104-019-36 | Sequence 37, Appl 1 |
| 27 | 29 | 63.0 | 199 | 9 | US-10-104-019-37 | Sequence 38, Appl 1 |
| 28 | 29 | 63.0 | 386 | 10 | US-09-737-178-127 | Sequence 127, App |
| 29 | 29 | 63.0 | 692 | 9 | US-10-101-464A-897 | Sequence 897, App |
| 30 | 29 | 63.0 | 732 | 10 | US-09-737-178-124 | Sequence 124, App |
| 31 | 29 | 63.0 | 804 | 9 | US-10-101-464A-890 | Sequence 890, App |
| 32 | 29 | 63.0 | 811 | 9 | US-09-992-598-57 | Sequence 57, Appl 1 |
| 33 | 29 | 63.0 | 811 | 9 | US-09-989-735-57 | Sequence 57, Appl 1 |
| 34 | 29 | 63.0 | 811 | 9 | US-09-989-735-57 | Sequence 57, Appl 1 |
| 35 | 29 | 63.0 | 811 | 9 | US-09-990-444-57 | Sequence 57, Appl 1 |
| 36 | 29 | 63.0 | 811 | 9 | US-09-989-730-57 | Sequence 57, Appl 1 |
| 37 | 29 | 63.0 | 811 | 9 | US-09-990-436-57 | Sequence 57, Appl 1 |
| 38 | 29 | 63.0 | 811 | 9 | US-09-991-181-57 | Sequence 57, Appl 1 |
| 39 | 29 | 63.0 | 811 | 9 | US-09-989-734-57 | Sequence 57, Appl 1 |
| 40 | 29 | 63.0 | 811 | 9 | US-09-997-653-57 | Sequence 57, Appl 1 |
| 41 | 29 | 63.0 | 811 | 9 | US-10-174-590-414 | Sequence 414, App |
| 42 | 29 | 63.0 | 811 | 9 | US-10-176-758-414 | Sequence 414, App |
| 43 | 29 | 63.0 | 811 | 9 | US-10-175-737-414 | Sequence 414, App |
| 44 | 29 | 63.0 | 811 | 9 | US-09-993-667-57 | Sequence 57, Appl 1 |
| 45 | 29 | 63.0 | 811 | 9 | US-09-993-667-57 | Sequence 57, Appl 1 |

ALIGNMENTS

RESULT 1
US-10-001-843-127
; Sequence 127, Application US/10001843
; Patent No. US20020132255A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Reclon, Heave
; APPLICANT: Cafferkey, Robert
; APPLICANT: Liu, Yongming
; APPLICANT: Liu, Chonghua
; APPLICANT: Turner, Leah
; TITLE OR INVENTION: Compositions and Methods Relating to Breast Specific Genes and
; FILE REFERENCE: DRX-0267
; CURRENT APPLICATION NUMBER: US/10/001.843
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,992
; PRIOR FILING DATE: 2000-11-20
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 127
; LENGTH: 40
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-001-843-127

Query Match 69.6%; Score 32; DB 12; Length 40;
Best Local Similarity 62.5%; Pred. No. 6.6;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
I I I I I I I I
Db 18 NFRYTNFS 25

RESULT 2
US-10-003-356-8
; Sequence 8, Application US/10003356
; Patent No. US20020146418A1
; GENERAL INFORMATION:
; APPLICANT: Lok, Si
; APPLICANT: Holloway, James L.

```

; TITLE OF INVENTION: Human V2 Vomeronasal Receptor
; FILE REFERENCE: 00-107
; CURRENT APPLICATION NUMBER: US/10/003,356
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: 60/252,373
; PRIOR FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 927
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Chimeric receptor.
US-10-003-356-8

Query Match
Best Local Similarity 69.6%; Score 32; DB 12; Length 927;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNF 7
    ||| |||
Db 523 NFRSTNF 529

RESULT 3
US-09-402-100-4
; Sequence 4, Application US/09402100
; Patent No. US20010019834A1
; GENERAL INFORMATION:
; APPLICANT: Daewoong Pharmaceutical Co, LTD
; APPLICANT: Kim, Byung-O
; APPLICANT: Shln, Sung-Seup
; APPLICANT: Yu, Young-Hyo
; APPLICANT: Park, Myung-Hwan
; APPLICANT: Choi, Deok-Uoon
; APPLICANT: Jung, Hyung-Jin
; TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Helicobacter
; FILE REFERENCE: 0136/06140
; CURRENT APPLICATION NUMBER: US/09/402,100
; CURRENT FILING DATE: 1999-09-27
; EARLIER APPLICATION NUMBER: KR 97-11950
; EARLIER FILING DATE: 1997-03-31
; EARLIER APPLICATION NUMBER: KR 97-11951
; EARLIER FILING DATE: 1997-03-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1338
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Caga/CTXA2B Chimeric protein
US-09-402-100-4

Query Match
Best Local Similarity 69.6%; Score 32; DB 10; Length 1338;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
    ||| |||
Db 505 NFRYTN 510

RESULT 4
US-09-864-761-45646
; Sequence 45646, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```

```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 45646
; LENGTH: 39
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011255.3
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.83
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.62
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.72
US-09-864-761-45646

Query Match
Best Local Similarity 67.4%; Score 31; DB 10; Length 39;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 FRYTNFA 8
    ||| |||
Db 7 FRYNNFA 13

RESULT 5
US-09-811-284-228
; Sequence 228, Application US/09811284
; Patent No. US20020058306A1
; GENERAL INFORMATION:
; APPLICANT: Vogeli, Gabriel
; TITLE OF INVENTION: No. US20020058306A1 G Protein-Coupled Receptors
; FILE REFERENCE: 001670U1
```

```

; CURRENT APPLICATION NUMBER: US/09/811,284
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: 60/189,783
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,907
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,918
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,960
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/189,917
; PRIOR FILING DATE: 2000-03-16
; PRIOR APPLICATION NUMBER: 60/192,945
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,916
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,923
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,933
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,830
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/192,234
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/192,155
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/192,935
; PRIOR FILING DATE: 2000-03-29
; NUMBER OF SEQ ID NOS: 258
; SOFTWARE: Patentln version 3.0
; SEQ ID NO 228
; LENGTH: 187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-811-284-228
```

```

Query Match          67.4%; Score 31; DB 10; Length 187;
Best Local Similarity 62.5%; Pred. No. 47;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 NFRYTNFA 8
    |||| : :|
Db 26 NFRYCSYA 33
```

```

RESULT 6
US-09-928-175-15
; Sequence 15, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 15
; LENGTH: 311
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-15
```

```

Query Match          67.4%; Score 31; DB 10; Length 311;
Best Local Similarity 62.5%; Pred. No. 78;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```

QY 1 NFRYTNFA 8
    |||| : :|
Db 278 NFRYCSYA 285
```

```

RESULT 7
US-09-928-175-10
; Sequence 10, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 10
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-10
```

```

Query Match          67.4%; Score 31; DB 10; Length 359;
Best Local Similarity 62.5%; Pred. No. 90;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 NFRYTNFA 8
    |||| : :|
Db 326 NFRYCSYA 333
```

```

RESULT 8
US-09-928-175-5
; Sequence 5, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-5
```

```

Query Match          67.4%; Score 31; DB 10; Length 383;
Best Local Similarity 62.5%; Pred. No. 96;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 NFRYTNFA 8
    |||| : :|
Db 350 NFRYCSYA 357
```

```

RESULT 9
US-09-928-175-13
```

```
; Sequence 13, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-13
```

```
Query Match          67.4%; Score 31; DB 10; Length 646;
Best Local Similarity 62.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 NFRYTNFA 8
      |||| : : |
Db      278 NFRYCSYA 285
```

```
RESULT 10
US-09-928-175-12
; Sequence 12, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 682
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-12
```

```
Query Match          67.4%; Score 31; DB 10; Length 682;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 NFRYTNFA 8
      |||| : : |
Db      314 NFRYCSYA 321
```

```
RESULT 11
US-09-928-175-8
; Sequence 8, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
```

```
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 694
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-8
```

```
Query Match          67.4%; Score 31; DB 10; Length 694;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 NFRYTNFA 8
      |||| : : |
Db      326 NFRYCSYA 333
```

```
RESULT 12
US-09-965-536A-6
; Sequence 6, Application US/09965536A
; Publication No. US20030027323A1
; GENERAL INFORMATION:
; APPLICANT: FEDER, J. N.
; APPLICANT: MINTIER, G.
; APPLICANT: RAMANATHAN, C. S.
; APPLICANT: HAWKEN, D. R.
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPBRM5,
; TITLE OF INVENTION: EXPRESSED HIGHLY IN BRAIN AND OVARIAN TISSUES
; FILE REFERENCE: D0041NP
; CURRENT APPLICATION NUMBER: US/09/965,536A
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,713
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/261,781
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/306,605
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/310,436
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-536A-6
```

```
Query Match          67.4%; Score 31; DB 9; Length 713;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
```

```
OY      1 NFRYTNFA 8
      |||| : : |
Db      345 NFRYCSYA 352
```

```
RESULT 13
US-09-928-175-3
; Sequence 3, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszty, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-Rich G Protein Coupled Receptor-8 Molecules and
```

```
; TITLE OF INVENTION: Uses thereof
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 718
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-3

Query Match          67.4%; Score 31; DB 10; Length 718;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
DB 350 NFRYCSYA 357

RESULT 14
US-09-928-175-7
; Sequence 7, Application US/09928175
; Patent No. US20020123618A1
; GENERAL INFORMATION:
; APPLICANT: Paszly, Christopher J.
; APPLICANT: Gong, Jianhua
; APPLICANT: Daugherty, Betsy
; APPLICANT: Rogers, No. US20020123618A1ma
; TITLE OF INVENTION: Leucine-rich G Protein Coupled Receptor-8 Molecules and
; FILE REFERENCE: 00-1229
; CURRENT APPLICATION NUMBER: US/09/928,175
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/224,455
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 730
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-928-175-7

Query Match          67.4%; Score 31; DB 10; Length 730;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
DB 362 NFRYCSYA 369

RESULT 15
US-09-965-536A-2
; Sequence 2, Application US/09965536A
; Publication No. US20030027323A1
; GENERAL INFORMATION:
; APPLICANT: FEDER, J. N.
; APPLICANT: MINTIER, G.
; APPLICANT: RAMANATHAN, C. S.
; APPLICANT: HAWKEN, D. R.
; TITLE OF INVENTION: A NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR, HGPRBMV5,
; FILE REFERENCE: D0041NP
; CURRENT APPLICATION NUMBER: US/09/965,536A
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: 60/235,713
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/261,781
```

```
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: 60/306,605
; PRIOR FILING DATE: 2001-07-19
; PRIOR APPLICATION NUMBER: 60/310,436
; PRIOR FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 737
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-536A-2

Query Match          67.4%; Score 31; DB 9; Length 737;
Best Local Similarity 62.5%; Pred. No. 1.8e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
DB 369 NFRYCSYA 376
```

Search completed: March 25, 2003, 08:23:42
Job time : 10.4545 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:21:42 ; Search time 9.54545 Seconds
(without alignments)
27.742 Million cell updates/sec

Title: US-09-646-532B-3

Perfect score: 46

Sequence: 1 NFRYTNEFX 9

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTC10S.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|---------------------|--------------------|
| 1 | 38 | 82.6 | 28 | 6 5521285-3 | Patent No. 5521285 |
| 2 | 32 | 69.6 | 245 | 4 US-09-502-653-12 | Sequence 12, Appl |
| 3 | 32 | 69.6 | 859 | 1 US-08-053-614-2 | Sequence 2, Appl |
| 4 | 32 | 69.6 | 859 | 1 US-08-316-397B-2 | Sequence 2, Appl |
| 5 | 32 | 69.6 | 859 | 2 US-09-034-306-2 | Sequence 2, Appl |
| 6 | 32 | 69.6 | 859 | 4 US-09-259-437-2 | Sequence 2, Appl |
| 7 | 32 | 69.6 | 859 | 5 PCT-US93-09782-2 | Sequence 2, Appl |
| 8 | 32 | 69.6 | 1147 | 3 US-08-470-260-5 | Sequence 5, Appl |
| 9 | 32 | 69.6 | 1147 | 4 US-08-471-491-5 | Sequence 5, Appl |
| 10 | 32 | 69.6 | 1147 | 4 US-08-466-662-5 | Sequence 5, Appl |
| 11 | 32 | 69.6 | 1181 | 1 US-08-053-614-4 | Sequence 4, Appl |
| 12 | 32 | 69.6 | 1181 | 1 US-08-316-397B-4 | Sequence 4, Appl |
| 13 | 32 | 69.6 | 1181 | 2 US-09-034-306-4 | Sequence 4, Appl |
| 14 | 32 | 69.6 | 1181 | 4 US-09-259-437-4 | Sequence 4, Appl |
| 15 | 32 | 69.6 | 1181 | 5 PCT-US93-09782-4 | Sequence 4, Appl |
| 16 | 32 | 69.6 | 3289 | 2 US-08-477-451-2 | Sequence 2, Appl |
| 17 | 31 | 67.4 | 9 | 1 US-08-189-772-3 | Sequence 3, Appl |
| 18 | 31 | 67.4 | 9 | 1 US-08-188-277B-12 | Sequence 12, Appl |
| 19 | 31 | 67.4 | 377 | 1 US-08-188-277B-2 | Sequence 2, Appl |
| 20 | 31 | 67.4 | 377 | 2 US-08-429-964-78 | Sequence 78, Appl |
| 21 | 31 | 67.4 | 377 | 2 US-08-429-964-80 | Sequence 80, Appl |
| 22 | 31 | 67.4 | 377 | 2 US-08-429-964-80 | Sequence 80, Appl |
| 23 | 31 | 67.4 | 630 | 3 US-08-771-986A-2 | Sequence 2, Appl |
| 24 | 31 | 67.4 | 630 | 3 US-08-769-802A-2 | Sequence 2, Appl |
| 25 | 31 | 67.4 | 823 | 1 US-08-461-551-2 | Sequence 2, Appl |
| 26 | 31 | 67.4 | 823 | 1 US-09-037-621A-2 | Sequence 2, Appl |
| 27 | 31 | 67.4 | 1030 | 4 US-09-091-117-2 | Sequence 2, Appl |

| | | | | | |
|----|----|------|------|---------------------|-------------------|
| 28 | 31 | 67.4 | 1163 | 1 US-08-239-474A-11 | Sequence 11, Appl |
| 29 | 31 | 67.4 | 1163 | 2 US-08-732-495-11 | Sequence 11, Appl |
| 30 | 31 | 67.4 | 1163 | 4 US-09-178-252-21 | Sequence 21, Appl |
| 31 | 31 | 67.4 | 1186 | 1 US-08-602-737-6 | Sequence 6, Appl |
| 32 | 31 | 67.4 | 1186 | 4 US-09-001-982-6 | Sequence 6, Appl |
| 33 | 31 | 67.4 | 1189 | 1 US-07-828-788A-16 | Sequence 16, Appl |
| 34 | 31 | 67.4 | 1189 | 1 US-08-356-034-6 | Sequence 6, Appl |
| 35 | 31 | 67.4 | 1189 | 1 US-08-602-737-2 | Sequence 2, Appl |
| 36 | 31 | 67.4 | 1189 | 2 US-08-980-071-2 | Sequence 2, Appl |
| 37 | 31 | 67.4 | 1189 | 2 US-08-980-071-6 | Sequence 6, Appl |
| 38 | 31 | 67.4 | 1189 | 2 US-08-980-071-6 | Sequence 6, Appl |
| 39 | 31 | 67.4 | 1189 | 2 US-08-980-071-8 | Sequence 8, Appl |
| 40 | 31 | 67.4 | 1189 | 2 US-08-980-071-10 | Sequence 10, Appl |
| 41 | 31 | 67.4 | 1189 | 2 US-08-980-071-12 | Sequence 12, Appl |
| 42 | 31 | 67.4 | 1189 | 2 US-08-980-071-59 | Sequence 59, Appl |
| 43 | 31 | 67.4 | 1189 | 2 US-08-980-071-61 | Sequence 61, Appl |
| 44 | 31 | 67.4 | 1189 | 2 US-08-757-536-2 | Sequence 2, Appl |
| 45 | 31 | 67.4 | 1189 | 2 US-08-757-536-4 | Sequence 4, Appl |

ALIGNMENTS

```

RESULT 1
5521285-3
; Patent No. 5521285
; APPLICANT: POMATO, NICHOLAS, BOS, EBO S.; RANSOM,
; JANET H.; HANNA, MICHAEL G. JR.
; TITLE OF INVENTION: CTA 28A32, THE ANTIGEN RECOGNIZED BY
; MCA 28A32
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/41,529
; FILING DATE: 01-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 508,373
; FILING DATE: 12-APR-1990
; APPLICATION NUMBER: 38,811
; FILING DATE: 15-APR-1987
; APPLICATION NUMBER: 697,078
; FILING DATE: 31-JAN-1985
; APPLICATION NUMBER: 575,533
; FILING DATE: 31-JAN-1984
; SEQ ID NO:3:
; LENGTH: 28
5521285-3

Query Match      82.6%; Score 38; DB 6; Length 28;
Best Local Similarity 87.5%; Pred. No. 0.44;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 NFRYTNEFX 8
      11111111
DB      15 NFRYTNEFX 22

RESULT 2
US-09-502-653-12
; Sequence 12, Application US/09502653
; Patent No. 6331426
; GENERAL INFORMATION:
; APPLICANT: Bj rrvad, Mads Eskelund
; APPLICANT: Clausen, Ib Groth
; APPLICANT: Schlein, Martin
; APPLICANT: Bech, Lisbeth
; APPLICANT: stergaard, Peter Rahbek
; APPLICANT: Sj holm, Carsten
; TITLE OF INVENTION: NOVEL GALACTANASRS
; FILE REFERENCE: 5481.200-US
; CURRENT APPLICATION NUMBER: US/09/502,653
; CURRENT FILING DATE: 2000-02-11
; EARLIER APPLICATION NUMBER: PA 1999 00184
; EARLIER FILING DATE: 1999-02-11

```

EARLIER APPLICATION NUMBER: PA 1999 00799
EARLIER FILING DATE: 1999-06-07
EARLIER APPLICATION NUMBER: 60/125,885
EARLIER FILING DATE: 1999-03-24
EARLIER APPLICATION NUMBER: 60/138,445
EARLIER FILING DATE: 1999-06-10
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 12
LENGTH: 245
TYPE: PRT
ORGANISM: Bacillus agaradhaerens AC13 (DSM 8721)
US-09-502-653-12

Query Match 69.6%; Score 32; DB 4; Length 245;
Best Local Similarity 71.4%; Pred. No. 56;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTN 7
||:|:
Db 47 NFRYSNF 53

RESULT 3
US-08-053-614-2
Sequence 2, Application US/08053614
Patent No. 5403924
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMMURU, MURALI K. R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P. C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-053-614-2

Query Match 69.6%; Score 32; DB 1; Length 859;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
||:|:
Db 505 NFRYTN 510

RESULT 4
US-08-316-397B-2
Sequence 2, Application US/08316397B
Patent No. 573740
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TUMMURU, MURALI K. R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P. C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-2

Query Match 69.6%; Score 32; DB 1; Length 859;
Best Local Similarity 83.3%; Pred. No. 2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
||:|:
Db 505 NFRYTN 510

RESULT 5
US-09-034-306-2
Sequence 2, Application US/09034306
Patent No. 5876943
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOUS
APPLICANT: TUMMURU, MURALI K. R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P. C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:


```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-2

Query Match
Best Local Similarity 69.6%; Score 32; DB 2; Length 859;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NFRYTN 6
Db 505 NFKYTN 510

RESULT 6
US-09-259-437-2
Sequence 2, Application US/09259437
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,437
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

```

```

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-259-437-2

Query Match
Best Local Similarity 69.6%; Score 32; DB 4; Length 859;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NFRYTN 6
Db 505 NFKYTN 510

RESULT 7
PCT-US93-09782-2
Sequence 2, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 859 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-2

Query Match
Best Local Similarity 69.6%; Score 32; DB 5; Length 859;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 NFRYTN 6
Db 505 NFKYTN 510

RESULT 8
US-08-470-260-5
Sequence 5, Application US/08470260
Patent No. 6077706
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello

```

APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: USA
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/470,260
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/256,848
FILING DATE: 21-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: McClung, Barbara G.
REGISTRATION NUMBER: 33,113
REFERENCE/DOCKET NUMBER: 0316.001
TELEPHONE: (510) 601-2708
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1147 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-470-260-5

Query Match 69.6%; Score 32; DB 3; Length 1147;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
||:||||
DB 505 NFRYTN 510

RESULT 9
US-08-471-491-5
Sequence 5, Application US/08471491B
Patent No. 6090611
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0044
CURRENT APPLICATION NUMBER: US/08/471,491B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-08-471-491-5

Query Match 69.6%; Score 32; DB 3; Length 1147;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
||:||||
DB 505 NFRYTN 510

RESULT 10
US-08-466-662-5
Sequence 5, Application US/08466662B
Patent No. 6130059
GENERAL INFORMATION:
APPLICANT: Covacci, Antonello
APPLICANT: Bugnoli, Massimo
APPLICANT: Telford, John
APPLICANT: Macchia, Giovanni
APPLICANT: Rappuoli, Rino
TITLE OF INVENTION: Helicobacter Pylori Proteins Useful For Vaccines And
FILE REFERENCE: CHIR0057
CURRENT APPLICATION NUMBER: US/08/466,662B
CURRENT FILING DATE: 1995-06-06
NUMBER OF SEQ ID NOS: 8
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 5
LENGTH: 1147
TYPE: PRT
ORGANISM: Helicobacter pylori
US-08-466-662-5

Query Match 69.6%; Score 32; DB 4; Length 1147;
Best Local Similarity 83.3%; Pred. No. 2.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
||:||||
DB 505 NFRYTN 510

RESULT 11
US-08-053-614-4
Sequence 4, Application US/08053614
Patent No. 5403924
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TUMWURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/053,614
FILING DATE: 19930426
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009

TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-053-614-4

Query Match
Best Local Similarity 83.3%; Score 32; DB 1; Length 1181;
Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
11:111
DB 505 NFRYTN 510

RESULT 12
US-08-316-397B-4
Sequence 4, Application US/08316397B
Patent No. 5733740
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
CORRESPONDENCE ADDRESS:
ADDRESS: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-316-397B-4

Query Match
Best Local Similarity 83.3%; Score 32; DB 1; Length 1181;
Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
11:111
DB 505 NFRYTN 510

RESULT 13

US-09-034-306-4
Sequence 4, Application US/09034306
Patent No. 5876943
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
CORRESPONDENCE ADDRESS:
ADDRESS: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/034,306
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-034-306-4

Query Match
Best Local Similarity 83.3%; Score 32; DB 2; Length 1181;
Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
11:111
DB 505 NFRYTN 510

RESULT 14
US-09-259-437-4
Sequence 4, Application US/09259437
Patent No. 6153390
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: HARRY KLEANTHOS
APPLICANT: TUMMURU, MURALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
NUMBER OF SEQUENCES: 4
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION AND GASTRIC CARCINOMA
CORRESPONDENCE ADDRESS:
ADDRESS: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/259,437
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/316,397
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.030
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-259-437-4

Query Match 69.6%; Score 32; DB 4; Length 1181;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
11:111
DB 505 NFRYTN 510

RESULT 15
PCT-US93-09782-4
Sequence 4, Application PC/TUS9309782
GENERAL INFORMATION:
APPLICANT: COVER, TIMOTHY L.
APPLICANT: BLASER, MARTIN J.
APPLICANT: TOMMORO, MORALI K.R.
TITLE OF INVENTION: THE TAGA GENE AND METHODS FOR DETECTING
TITLE OF INVENTION: PREDISPOSITION TO PEPTIC ULCERATION
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/09782
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: SPRATT, GWENDOLYN D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 2200.009
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1181 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-09782-4

Query Match 69.6%; Score 32; DB 5; Length 1181;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTN 6
11:111
DB 505 NFRYTN 510

Search completed: March 25, 2003, 08:23:01
Job time: 11.5455 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:20:22 ; Search time 9.81818 Seconds
(without alignments)
88.123 Million cell updates/sec

Title: US-09-646-532b-3

Perfect score: 46

Sequence: 1 NFRYTNFA 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: PIR.73:*
2: PIR1:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|--------|--------------------|
| 1 | 34 | 73.9 | 222 | D82132 | hypothetical prote |
| 2 | 34 | 73.9 | 2802 | F97686 | cyclic beta-(1-2) |
| 3 | 34 | 73.9 | 2831 | A12911 | beta (1->2) glucan |
| 4 | 33 | 71.7 | 104 | H97589 | hypothetical prote |
| 5 | 33 | 71.7 | 104 | AG2811 | hypothetical prote |
| 6 | 33 | 71.7 | 157 | T17883 | major capsid prote |
| 7 | 33 | 71.7 | 187 | H61347 | probable decarboxy |
| 8 | 33 | 71.7 | 262 | T22371 | hypothetical prote |
| 9 | 33 | 71.7 | 437 | VCXECV | major capsid prote |
| 10 | 33 | 71.7 | 511 | F70110 | conserved hypotet |
| 11 | 33 | 71.7 | 520 | T18124 | probable capsid pr |
| 12 | 33 | 71.7 | 547 | T44743 | probable thiamin b |
| 13 | 33 | 71.7 | 1315 | T41055 | membrane thiamin b |
| 14 | 32 | 66.6 | 115 | S22588 | VI protein - tomat |
| 15 | 32 | 66.6 | 258 | T25361 | hypothetical prote |
| 16 | 32 | 66.6 | 315 | A36944 | outer membrane pro |
| 17 | 32 | 66.6 | 346 | C90590 | hypothetical prote |
| 18 | 32 | 66.6 | 369 | E86307 | Similar to polygal |
| 19 | 32 | 66.6 | 467 | T02238 | glucosyl transfera |
| 20 | 32 | 66.6 | 595 | T06111 | hypothetical prote |
| 21 | 32 | 66.6 | 604 | T08302 | hypothetical prote |
| 22 | 32 | 66.6 | 604 | T08222 | hypothetical prote |
| 23 | 32 | 66.6 | 607 | B75614 | thiamin biosynthes |
| 24 | 32 | 66.6 | 713 | T20090 | hypothetical prote |
| 25 | 32 | 66.6 | 1151 | S48431 | probable membrane |
| 26 | 32 | 66.6 | 1167 | B71924 | cag island protein |
| 27 | 32 | 66.6 | 1186 | C64588 | cag pathogenicity |
| 28 | 32 | 66.6 | 1215 | B48281 | cytotoxin-associat |
| 29 | 32 | 66.6 | 1465 | T23056 | hypothetical prote |

| | | | | | | |
|----|----|------|------|---|--------|---------------------|
| 30 | 31 | 67.4 | 145 | 2 | D90358 | hypothetical prote |
| 31 | 31 | 67.4 | 155 | 2 | F70239 | transposase-like p |
| 32 | 31 | 67.4 | 211 | 2 | D96996 | uncharacterized lo |
| 33 | 31 | 67.4 | 300 | 2 | S25706 | Zf-cad1 protein - |
| 34 | 31 | 67.4 | 334 | 2 | G90492 | hypothetical prote |
| 35 | 31 | 67.4 | 377 | 2 | A53044 | geranylgeranyl-dip |
| 36 | 31 | 67.4 | 377 | 2 | B53044 | geranylgeranyl-dip |
| 37 | 31 | 67.4 | 379 | 2 | T22392 | hypothetical prote |
| 38 | 31 | 67.4 | 437 | 2 | G70019 | conserved hypotet |
| 39 | 31 | 67.4 | 655 | 2 | JC7140 | protoxin - Bacillu |
| 40 | 31 | 67.4 | 662 | 2 | E97973 | hypothetical prote |
| 41 | 31 | 67.4 | 819 | 2 | T07319 | hypothetical prote |
| 42 | 31 | 67.4 | 823 | 2 | S04181 | parasporeal crystal |
| 43 | 31 | 67.4 | 856 | 2 | D86258 | protein F5011.10 f |
| 44 | 31 | 67.4 | 1030 | 2 | S73944 | hypothetical prote |
| 45 | 31 | 67.4 | 1189 | 2 | S00944 | parasporeal crystal |

ALIGNMENTS

RESULT 1
D82132
hypothetical protein VC1981 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence-revision 20-Aug-2000 #text-change 02-Feb-2001
C:Accession: D82132
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.
Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers
I, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A>Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; M01D:20406833; PMID:10952301
A:Accession: D82132
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-222 <HEI>
A:Cross-references: GB:AE004273; GB:AE003852; NID:9656517; PIDN:AAF95129.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1981
A:Map position: 1

Query Match
Best Local Similarity 73.9%; Score 34; DB 2; Length 222;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
II :|||
DB 84 NFAHTNFA 91

RESULT 2
F97686
cyclic beta-(1-2) glucan synthetase [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #text-change 11-Jan-2002
C:Accession: F97686
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: F97686
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2802 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK88447.1; PID:q15157946; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_4949
A:Map position: circular chromosome

Query Match 73.9%; Score 34; DB 2; Length 2802;

Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
|:|||||
Db 1368 NYQYTNF 1374

RESULT 3

AI2911

beta (1->2) glucan biosynthesis protein [imported] - Agrobacterium tumefaciens (strain C

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AI2911

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCell

: Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AI2911

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2831 <KUR>

A:Cross-references: GB:AE008688; PIDN:AL43711.1; PID:917741241; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: chvB

A:Map position: circular chromosome

Query Match 73.9%; Score 34; DB 2; Length 2831;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
|:|||||
Db 1397 NYQYTNF 1403

RESULT 4

H97589

hypothetical protein AGR_C3504 [imported] - Agrobacterium tumefaciens (strain C58, Cere

C:Species: Agrobacterium tumefaciens

C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: H97589

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: H97589

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-104 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK87673.1; PID:915157030; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C3504

A:Map position: circular chromosome

Query Match 71.7%; Score 33; DB 2; Length 104;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
|:|||||
Db 5 NFRYTHY 11

RESULT 5

AG2811

hypothetical protein Atul913 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AG2811

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl

: Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AG2811

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-104 <KUR>

A:Cross-references: GB:AE008688; PIDN:AL42909.1; PID:917740364; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atul913

A:Map position: circular chromosome

Query Match 71.7%; Score 33; DB 2; Length 104;
Best Local Similarity 71.4%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
|:|||||
Db 5 NFRYTHY 11

RESULT 6

117883

major capsid protein homolog A382R - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: 117883

R:Graves, M.V.; Van Etten, J.L.

submitted to the EMBL data library, May 1999

A:Reference number: Z18806

A:Accession: 117883

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-157 <GRN>

A:Cross-references: EMBL:U42580; NID:94028896; PIDN:ACG6750.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:

A:Note: A382R

Query Match 71.7%; Score 33; DB 2; Length 157;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RYTNA 8
|:|||||
Db 125 RYTNA 130

RESULT 7

H81347

probable decarboxylase Cj0768c [imported] - Campylobacter jejuni (strain NCTC 11168)

C:Species: Campylobacter jejuni

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002

C:Accession: H81347

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Bar

Nature 403, 665-668, 2000

A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A:Reference number: AB1250; MUID:20150912; PMID:10688204

A:Accession: H81347

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-187 <PAR>

A:Cross-references: GB:AL139076; GB:AL111168; NID:96968128; PIDN:CAB73033.1; PID:9696

A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0768c
 C:Superfamily: dedf protein

Query Match 71.7%; Score 33; DB 2; Length 187;
 Best Local Similarity 71.4%; Pred. No. 21;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTNF 7
 ||:|:|:|
 DB 56 NFRYTNF 62

RESULT 8

hypothetical protein F47H4.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22371

R:Steward, C.
 Submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19535
 A:Accession: T22371
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-262 <MIL>
 A:Cross-references: EMBL:Z292786; PIDDN:CAB07207.1; GSPDB:GN00023; CESP:F47H4.8
 C:Genetics:
 A:Experimental source: clone F47H4
 A:Gene: CESP:F47H4.8
 A:Map position: 5
 A:Introns: 67/2; 98/2; 131/2; 186/1; 199/3

Query Match 71.7%; Score 33; DB 2; Length 262;
 Best Local Similarity 71.4%; Pred. No. 29;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 NFRYTNF 7
 ||:|:|:|
 DB 206 NFRYTNF 212

RESULT 9

major capsid protein - Chlorella virus PBCV-1
 C:Species: Chlorella virus PBCV-1
 C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 29-Oct-1999
 C:Accession: A40245; T17933

R:Graves, M.V.; Meints, R.H.
 Virology 188, 198-207, 1992
 A:Title: Characterization of the major capsid protein and cloning of its gene from algal
 A:Reference number: A40245; MUID:92230218; PMID:1566573
 A:Accession: A40245
 A:Molecule type: DNA
 A:Residues: 1-437 <GRA>

A:Cross-references: GB:M85052; NID:93232323; PIDDN:AAA88828.1; PID:93232324
 R:Graves, M.V.; Van Etten, J.L.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z18806
 A:Accession: T17933

A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-437 <GR2>
 A:Cross-references: EMBL:U42580; NID:94028896; PIDDN:AAC96798.1
 A:Experimental source: specific host Chlorella strain NC64A
 C:Genetics:
 A:Note: A430L

C:Superfamily: T1pula Iridescent virus major capsid protein
 C:Keywords: coat protein; glycoprotein
 F:47,189,370,376/Binding site: carbohydrate (Asn) (covalent) #status predicted
 Query Match 71.7%; Score 33; DB 1; Length 437;

Best Local Similarity 100.0%; Pred. No. 49;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RYTNFA 8
 |||||
 DB 33 RYTNFA 38

RESULT 10

conserved hypothetical protein BB0086 - Lyme disease spirochete
 C:Species: Borrelia burgdorferi (Lyme disease spirochete)
 C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 29-Sep-1999
 C:Accession: F70110

R:Praser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Latilga, R.; Wh
 son, D.; Peterson, J.; Kerville, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu
 ; Bowman, C.; Garland, S.; Fujil, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
 Nature 390, 580-586, 1997
 A:Authors: Smith, H.O.; Venter, J.C.

A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
 A:Reference number: A70100; MUID:98065943; PMID:9403685
 A:Accession: F70110
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-511 <KLE>

A:Cross-references: GB:AE001121; GB:AE000783; NID:92687963; PIDDN:AAC66471.1; PID:9268
 A:Experimental source: strain B31
 C:Superfamily: conserved hypothetical protein H1117

Query Match 71.7%; Score 33; DB 2; Length 511;
 Best Local Similarity 75.0%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 NFRYTNFA 8
 |||||
 DB 423 NFRYTNFA 430

RESULT 11

probable capsid protein A622L - Chlorella virus PBCV-1
 C:Species: Chlorella virus PBCV-1
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T18124

R:Graves, M.V.; Van Etten, J.L.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z18806
 A:Accession: T18124
 A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-520 <GRA>
 A:Cross-references: EMBL:U42580; NID:94028896; PIDDN:AAC96953.1
 A:Experimental source: specific host Chlorella strain NC64A
 C:Genetics:
 A:Note: A622L

C:Keywords: coat protein

Query Match 71.7%; Score 33; DB 2; Length 520;
 Best Local Similarity 100.0%; Pred. No. 59;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RYTNFA 8
 |||||
 DB 33 RYTNFA 38

RESULT 12

probable thiamin biosynthesis protein thlC [imported] - Mycobacterium leprae
 C:Species: Mycobacterium leprae
 C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 18-Feb-2000
 C:Accession: T44743

R:James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, January 1998

A:Reference number: Z22831

A:Accession: T44743

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-547 <JAM>

A:Cross-references: EMBL:AL035159; PIDN:CAA22712.1

A:Experimental source: cosmid B1450

C:Genetics:

A:Gene: thtC

C:Superfamily: thiamin biosynthesis protein thtC

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 547;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNFA 8

Db 281 SFLYTNPFA 288

RESULT 13

membrane atpase - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999

C:Accession: T41055

R:Hubbert, H.; Duesterhoeft, A.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z21967

A:Accession: T41055

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1315 <HLL>

A:Cross-references: EMBL:AL031324; PIDN:CAA20449.1; GSPDB:GN00068; SPDB:SPCC1672.11C

A:Experimental source: strain 972n-; cosmid c1672

C:Genetics:

A:Gene: SPDB:SPCC1672.11C

A:Map position: 3

Query Match

Best Local Similarity 71.7%; Score 33; DB 2; Length 1315;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNF 7

Db 256 NFRYVNY 262

RESULT 14

S22588

V1 protein - tomato yellow leaf curl virus

C:Species: tomato yellow leaf curl virus

C:Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 20-Sep-1999

C:Accession: S22588; S39233; S39209

R:Kheyr-Pour, A.; Bendahmane, M.; Matzeit, V.; Accotto, G.P.; Crespi, S.; Gronenborn, B.

Nucleic Acids Res. 19, 6763-6769, 1991

A:Title: Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite

A:Reference number: S22588; MUID:92107660; PMID:1840676

A:Accession: S22588

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-115 <KHE>

A:Cross-references: EMBL:X61153; NID:962211; PIDN:CAA43462.1; PID:962212

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1991

R:Crespi, S.; Norris, E.; Vaira, A.; Bosco, D.; Accotto, G.

Submitted to the EMBL Data Library, December 1993

A:Description: A cloned DNA from a TYLCV isolate from Sicily showing low infectivity.

A:Reference number: S39233

A:Accession: S39233

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-115 <CRE>

A:Cross-references: EMBL:Z28390; NID:91041671; PIDN:CAA82235.1; PID:9433850

R:Norris, E.; Hidalgo, E.; Accotto, G.; Moriones, E.

submitted to the EMBL Data Library, August 1993

A:Description: High similarity among the tomato yellow leaf curl virus isolates from

A:Reference number: S39209

A:Accession: S39209

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-115 <NOR>

A:Cross-references: EMBL:Z25751; NID:9433655; PIDN:CAA81024.1; PID:9433656

C:Superfamily: tomato yellow leaf curl virus V2 protein

Query Match

Best Local Similarity 69.6%; Score 32; DB 2; Length 115;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNF 7

Db 58 NRRYTNF 64

RESULT 15

hypothetical protein T27E7.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T25361

R:Cummings, P.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z20023

A:Accession: T25361

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-258 <WLL>

A:Cross-references: EMBL:Z82284; PIDN:CAE05288.1; GSPDB:GN00022; CESP:T27E7.5

A:Experimental source: clone T27E7

C:Genetics:

A:Gene: CESP:T27E7.5

A:Map position: 4

A:Introns: 5/3; 33/1; 142/3; 192/3; 211/1

Query Match

Best Local Similarity 69.6%; Score 32; DB 2; Length 258;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6

Db 213 NRYTN 218

Search completed: March 25, 2003, 08:22:21
Job time : 11.8182 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:18:08 ; Search time 5.45455 Seconds
(without alignments)
68.436 Million cell updates/sec

Title: US-09-646-532B-3
Perfect score: 46
Sequence: 1 NFRYTNFAX 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|--------------------|
| 1 | 33 | 71.7 | 187 | 1 | Q9PFI1 campylobact |
| 2 | 33 | 71.7 | 436 | 1 | COAT_CHVP1 |
| 3 | 33 | 71.7 | 547 | 1 | THIC_MYCLE |
| 4 | 33 | 71.7 | 1315 | 1 | ATC9_SCHPO |
| 5 | 32 | 69.6 | 115 | 1 | YV1_TYLCM |
| 6 | 32 | 69.6 | 315 | 1 | OMP_ECOCI |
| 7 | 32 | 69.6 | 370 | 1 | Y205_HUMAN |
| 8 | 32 | 69.6 | 1147 | 1 | CGA1_HELPY |
| 9 | 32 | 69.6 | 1151 | 1 | ATC7_YEAST |
| 10 | 32 | 69.6 | 1167 | 1 | CAGA_HELPY |
| 11 | 32 | 69.6 | 1182 | 1 | CGA2_HELPY |
| 12 | 32 | 69.6 | 1186 | 1 | CAGA_HELPY |
| 13 | 31 | 67.4 | 377 | 1 | PGT1_HUMAN |
| 14 | 31 | 67.4 | 377 | 1 | PGT1_RAT |
| 15 | 31 | 67.4 | 753 | 1 | CKAA_HUMAN |
| 16 | 31 | 67.4 | 754 | 1 | LCR8_HUMAN |
| 17 | 31 | 67.4 | 819 | 1 | YCAI_CHLYD |
| 18 | 31 | 67.4 | 1030 | 1 | Y075_MYCPN |
| 19 | 31 | 67.4 | 1189 | 1 | CICA_BACTE |
| 20 | 31 | 67.4 | 1322 | 1 | SALI_MOUSE |
| 21 | 31 | 67.4 | 1324 | 1 | SALI_HUMAN |
| 22 | 31 | 67.4 | 1522 | 1 | CINI_LOBL |
| 23 | 30 | 65.2 | 146 | 1 | ANGI_PAPHA |
| 24 | 30 | 65.2 | 215 | 1 | 3F_DICDI |
| 25 | 30 | 65.2 | 322 | 1 | DIDH_RAT |
| 26 | 30 | 65.2 | 416 | 1 | PGK_MYCGE |
| 27 | 30 | 65.2 | 443 | 1 | CSN2_HUMAN |
| 28 | 30 | 65.2 | 621 | 1 | YPD7_CAEEL |
| 29 | 30 | 65.2 | 649 | 1 | YAT3_SCHPO |
| 30 | 30 | 65.2 | 717 | 1 | NAH4_RAT |
| 31 | 30 | 65.2 | 721 | 1 | RIR1_MYCGE |
| 32 | 30 | 65.2 | 721 | 1 | RIR1_MYCPN |
| 33 | 30 | 65.2 | 758 | 1 | PMT2_YEAST |

| | | | | | | |
|----|----|------|------|---|------------|--------------------|
| 34 | 30 | 65.2 | 1160 | 1 | CIDB_BACTU | Q45747 bacillus th |
| 35 | 30 | 65.2 | 1165 | 1 | CIDA_BACTA | P19415 bacillus th |
| 36 | 30 | 65.2 | 1171 | 1 | CIEA_BACTX | O57458 bacillus th |
| 37 | 29 | 63.0 | 154 | 1 | VE6_HPV07 | P36800 human papil |
| 38 | 29 | 63.0 | 154 | 1 | VE6_HPV40 | P36812 human papil |
| 39 | 29 | 63.0 | 199 | 1 | ATP6_CAEEL | P24888 caenorhabdi |
| 40 | 29 | 63.0 | 199 | 1 | YET4_HAEIN | O57213 haemophilus |
| 41 | 29 | 63.0 | 209 | 1 | SOMA_ANGTA | P08899 anguilla ja |
| 42 | 29 | 63.0 | 228 | 1 | YBEG_HAEIN | P44670 haemophilus |
| 43 | 29 | 63.0 | 237 | 1 | PSS_HELPY | O92m68 haemophilus |
| 44 | 29 | 63.0 | 237 | 1 | PSS_HELPY | O48269 haemophilus |
| 45 | 29 | 63.0 | 437 | 1 | AAC3_DICDI | P14197 dictyostell |

ALIGNMENTS

| RESULT 1 | ID | PAAD_CAMJE | STANDARD | PRT | 187 AA. |
|-----------------------|--|---------------|----------|-------------|---------|
| AC | Q9PFI1 | | | | |
| DT | 16-OCT-2001 (Rel. 40, Created) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last sequence update) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | |
| DE | Probable aromatic acid decarboxylase (EC 4.1.1.-). | | | | |
| GN | CJ0768C. | | | | |
| OS | Campylobacter jejuni. | | | | |
| OC | Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group; | | | | |
| OX | Campylobacter. | | | | |
| NCBI_TaxID=197; | | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-NCTC 11168; | | | | |
| RX | MEDLINE=20150912; PubMed=10688204; | | | | |
| RA | Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C., | | | | |
| RA | Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S., | | | | |
| RA | Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W., | | | | |
| RA | Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M., | | | | |
| RA | Whitehead S., Barrell B.G.; | | | | |
| RT | "The genome sequence of the food-borne pathogen Campylobacter jejuni | | | | |
| RT | reveals hypervariable sequences." | | | | |
| RL | Nature 403:665-668(2000). | | | | |
| CC | -I- SIMILARITY: BELONGS TO THE POLYPRENYL P-HYDROXYBENZOATE / | | | | |
| CC | PHENYLACETYLIC ACID DECARBOXYLASES FAMILY. | | | | |
| CC | | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@isb-sib.ch). | | | | |
| CC | | | | | |
| DR | EMBL: AL139076; CAB73033.1; | | | | |
| DR | InterPro: IPR003382; Flavoprotein. | | | | |
| DR | Pfam: PF02441; Flavoprotein; 1. | | | | |
| KW | Hypothetical protein; Lyase; Decarboxylase; Complete proteome. | | | | |
| SQ | SEQUENCE 187 AA; 20890 MW; 01D246C5EBCA7E7C CRC64; | | | | |
| Query Match | 71.7%; | Score 33; | DB 1; | Length 187; | |
| Best Local Similarity | 71.4%; | Pred No. 9.2; | | | |
| Matches | 5; | Conservative | 2; | Mismatches | 0; |
| | | | | Indels | 0; |
| | | | | Gaps | 0; |
| QY | 1 NFRYTNF 7 | | | | |
| | : : : | | | | |
| DB | 56 NFRYTNF 62 | | | | |
| RESULT 2 | | | | | |
| COAT_CHVP1 | STANDARD; | PRT; | 436 AA. | | |
| ID | COAT_CHVP1 | | | | |
| AC | P30328; | | | | |
| DT | 01-APR-1993 (Rel. 25, Created) | | | | |

```

DT 01-JUL-1993 (Rel. 26, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Major capsid protein (VP54).
GN A430L.
OS Paramexium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
RX MEDLINE=92230218; PubMed=1566573;
RA Graves M.V., Meints R.H.;
RT "Characterization of the major capsid protein and cloning of its gene
RL from algal virus PBCV-1."
RL Virology 188:198-207(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-Kb genome: map
RL positions 182 to 258."
RL Virology 223:303-317(1996).
CC -1 SUBUNIT: HOMODIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M85052; AAA88828.1; -
DR EMBL: U42580; AAC96798.1; -
DR PIR: A40245; VCXECV.
KM Coat protein; Glycoprotein; Late protein.
FT INT MET 0 0
FT DOMAIN 1 30 HYDROPHOBIC.
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 369 369 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 436 AA; 48034 MW; A71861ADBEF813B CXC64;

Query Match 71.7%; Score 33; DB 1; Length 436;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RYTNFA 8
    |||||
DB 32 RYTNFA 37

RESULT 3
THIC_MYCLE STANDARD: PRT: 547 AA.
AC Q9ZBL0;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Thiamine biosynthesis protein thic.
GN THIC OR M0294 OR MCB1450.28C.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RA MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,

```

```

RA Holroyd S., Hornsby T., Jagels K., Iacox C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1 FUNCTION: REQUIRED FOR THE SYNTHESIS OF THE HYDROMETHYLPIRIMIDINE
CC (HMP) MOIETY OF THIAMINE (4-AMINO-2-METHYL-5-
CC HYDROXYMETHYLPIRIMIDINE) (BY SIMILARITY).
CC -1 PATHWAY: Thiamine biosynthesis.
CC -1 SIMILARITY: BELONGS TO THE THIC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AL035159; CAA22712.1; -
DR EMBL: AL583918; CAC29802.1; -
DR Leptoma; M0294; -
DR InterPro; IPR002817; Thic.
DR Pfam; PF01964; Thic. 1.
DR ProDom; PD007048; Thic. 1.
DR TIGRFAMs; TIGR00190; thic. 1.
KW Thiamine biosynthesis; Complete proteome.
SQ SEQUENCE 547 AA; 59840 MW; 797DFB09DE462D58 CXC64;

Query Match 71.7%; Score 33; DB 1; Length 547;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNFA 8
    :|||
DB 281 SFLYTNFA 288

RESULT 4
ATC9_SCHPO STANDARD: PRT: 1315 AA.
ID ATC9_SCHPO
AC 074431;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE "Probable cation-transporting ATPase C1672.11C (EC 3.6.3.-).
GN SPC1672.11C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tiley A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volkart G., Aert R., Robben J., Grymoprez B.,
RA Weltjens I., Vansurels E., Klieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,

```

RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Mambutt R., Purrelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Motlier S.,
RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Patakhin J.,
RA Shporkovskiy G.V., Ussery D., Barrell B.G., Nurse P.,
RT "The genome sequence of Schizosaccharomyces pombe".
RL Nature 415:871-880(2002).
CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E2
CC ATPASES). SUBFAMILY V.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AL031324; CAA20449.1; -
DR InterPro: IPR001757; ATPase_E1-E2.
DR InterPro: IPR001454; Hlgase/hydriase.
DR Pfam: PF00122; E1-E2_ATPase_1.
DR Pfam: PF00702; Hydriase_1.
DR PRINTS: PR00119; CATATPASE.
DR PROSITE: PS00154; ATPase_E1-E2; 1.
KM Hypothetical protein: Transmembrane; Phosphorylation;
KV Magnesium; ATP-binding.
FT DOMAIN 1 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 161 POTENTIAL.
FT DOMAIN 162 167 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 168 190 POTENTIAL.
FT DOMAIN 191 324 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 325 347 POTENTIAL.
FT DOMAIN 348 350 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 351 369 POTENTIAL.
FT TRANSMEM 370 531 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 532 551 POTENTIAL.
FT DOMAIN 552 564 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 565 586 POTENTIAL.
FT DOMAIN 587 1072 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1073 1091 POTENTIAL.
FT DOMAIN 1092 1099 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1100 1117 POTENTIAL.
FT DOMAIN 1118 1135 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1136 1159 POTENTIAL.
FT DOMAIN 1160 1180 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1181 1203 POTENTIAL.
FT DOMAIN 1204 1216 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 1217 1236 POTENTIAL.
FT DOMAIN 1237 1253 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1254 1274 POTENTIAL.
FT DOMAIN 1275 1315 CYTOPLASMIC (POTENTIAL).
FT MOD_RES 619 619 PHOSPHORYLATION (BY SIMILARITY).
FT METAL 1015 1015 MAGNESIUM (BY SIMILARITY).
FT METAL 1019 1019 MAGNESIUM (BY SIMILARITY).
SQ SEQUENCE 1315 AA; 148786 MW; 081A50E59F2B0B5 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 1315;
Best Local Similarity 71.4%; Pred. No. 66;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NRRYTNF 7
|||||
DB 256 NRRYTNF 262

RESULT 5

VY1_TYLCM STANDARD: PRT: 115 AA.
ID VY1_TYLCM
AC P27270;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Hypothetical 13.3 kDa protein (V1 protein).
GN V1.
OS Tomato yellow leaf curl virus (strain Marnande) (TYLCV), and
OS Tomato yellow leaf curl virus (strain Murcia) (TYLCV).
OC Viruses; ssDNA viruses; Geminiviridae; Begomovirus.
OX NCBI_TaxID=10833, 37139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Marnande;
RX MEDLINE=9210760; PubMed=1840676;
RA Kheyri-Pour A., Bendahmane M., Matzelt V., Accotto G.P., Crespi S.,
RA Gronenborn B.;
RT "Tomato yellow leaf curl virus from sardina is a
RT whitefly-transmitted monopartite geminivirus.";
RL Nucleic Acids Res. 19:6763-6769(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Murcia;
RX MEDLINE=94256836; PubMed=8198442;
RA Norris E., Hidalgo E., Accotto G., Moriones E.;
RT "High similarity among the tomato yellow leaf curl virus isolates
RT from the west Mediterranean basin: the nucleotide sequence of an
RT infectious clone from Spain.";
RL Arch. Virol. 135:165-170(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

DR EMBL: K61153; CAA43462.1; -
DR EMBL: Z25751; CAA81024.1; -
DR PIR: S22588; S22588.
DR PIR: S39209; S39209.
DR PIR: S39233; S39233.
DR InterPro: IPR002511; Geminl_V1.
DR InterPro: IPR005159; WCCH.
DR Pfam: PF01524; Geminl_V1; 1.
DR Pfam: PF03716; WCCH; 1.
DR Prodom: PD002978; Geminl_V1; 1.
KV Hypothetical protein.
SQ SEQUENCE 115 AA; 13251 MW; D1AF6654C12A208 CRC64;
Query Match 69.6%; Score 32; DB 1; Length 115;
Best Local Similarity 85.7%; Pred. No. 8.9;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 NRRYTNF 7
|||||
DB 58 NRRYTNF 64
RESULT 6
OMP_ECOLI STANDARD: PRT: 315 AA.
ID OMP_ECOLI
AC P34210;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Outer membrane protease omp precursor (EC 3.4.23.-).
GN OMP OR OMPX.
OS Escherichia coli.
OC plasmid F.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A. AND SEQUENCE OF 24-43 AND 179-192.
RC STRAIN=K12 / ATCC 12435.
RX MEDLINE=94117370; PubMed=8288530;
RA Kaufmann A., Stierhof Y.-D., Henning U.,
RT "New outer membrane-associated protease of Escherichia coli K-12."
RL J. Bacteriol. 176:359-367(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / CR63.
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.,
RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes."
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
RN [3]
RP CHARACTERIZATION.
RC MEDLINE=20029733; PubMed=10561486;
RA Matsuo E., Sampei G., Mizobuchi K., Ito K.;
RT "The plasmid F Omp protease, a homologue of OmpT, as a potential
RT obstacle to E. coli-based protein production."
RL FEBS Lett. 461:6-8(1999).
CC -1- FUNCTION: PROTEASE: ALSO ACTS AS A RECEPTOR FOR BACTERIOPHAGE OX2.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A20.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X74278; CAA52338.1; -.
DR EMBL: AP001918; BAA97899.1; -.
DR PIR: S37473; S37473.
DR HSSP: P09169; 1178.
DR MEROPS: A26.002; -.
DR EcoGene: EG40042; ompP.
DR InterPro: IPR000036; OmpTln.
DR Pfam: PF01278; OmpTln; 1.
DR PRINTS: PR00482; OMPRTN.
DR PRODOM: PD011585; OmpTln; 1.
DR PROSITE: PS00834; OMPRTN_1; 1.
DR PROSITE: PS00835; OMPRTN_2; 1.
KM Hydrolyase; Aspartyl protease; Outer membrane; Transmembrane; Signal;
KM Phage recognition; Plasmid; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 315 OUTER MEMBRANE PROTEASE OMP.
FT ACT_SITE 103 103 BY SIMILARITY.
FT ACT_SITE 105 105 BY SIMILARITY.
FT ACT_SITE 230 230 BY SIMILARITY.
FT ACT_SITE 232 232 BY SIMILARITY.
SQ SEQUENCE 315 AA; 35499 MW; 45595973C6B2B149 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 315;
Best Local Similarity 71.4%; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

DE Hypothetical protein KIAA0205.
CN KIAA0205.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=97191544; PubMed=9039502;
RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y.,
RA Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. VI.
RT The coding sequences of 80 new genes (K1A0201-K1A0280) deduced by
RT analysis of cDNA clones from cell line KG-1 and brain."
RL DNA Res. 3:321-329(1996).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE 1-ACYL-SN-GLYCEROL-3-PHOSPHATE
CC ACYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: D86960; BAA13196.1; -.
DR InterPro: IPR002123; Acyltransferase.
KM Hypothetical protein; Phospholipid biosynthesis; Transferase;
KM Acyltransferase; Transmembrane.
FT TRANSMEM 22 42
FT TRANSMEM 342 362 POTENTIAL.
SQ SEQUENCE 370 AA; 43089 MW; 93498544EA651541 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 370;
Best Local Similarity 83.3%; Pred. No. 29;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 1 FRYTNE 7
DB 206 NRYRDNF 212

RESULT 7
Y205_HUMAN STANDARD; PRT; 370 AA.
AC Q92604;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

OY 2 FRYTNE 7
DB 132 FRYTNE 137

RESULT 8
CGAL_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
CN CGAL OR CAL.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N.,
RA Rappunli R.;
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal
RT ulcer."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=97121442; PubMed=8962108;

```

RA Censini S., Lange C., Xiang Z., Crabtree J., Ghara P.,
 RA Bordovsky M., Rappaport R., Covacci A.,
 RT "cagA, a pathogenicity island of Helicobacter pylori, encodes type I-
 RT specific and disease-associated virulence factors.",
 RT Proc. Natl. Acad. Sci. U.S.A. 93:14640-14653(1996).
 RN [3]
 RP SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506; 661-677;
 RP 900-914 AND 1062-1077.
 RA Hermann V., Hermann J., Rist M.,
 RA Submitted (Apr-1993) to the SWISS-PROT data bank.
 CC -1- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
 CC OR FUNCTION OF THE CYTOXIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X70039; CAA49633.1; -
 CC EMBL: AF282853; AAC44706.1; -
 CC InterPro: IPR005169; Caga.
 CC InterPro: IPR004355; IVSec_caga.
 CC Pfam: PF03507; Caga; 1.
 CC PRINTS: PR01553; TYPE4SSCAGA.
 CC KW Antigen.
 CC FT DOMAIN 880 885 POLY-ASN.
 CC FT CONFLICT 320 320 G -> A (IN REF. 3).
 CC FT CONFLICT 325 325 P -> F (IN REF. 3).
 CC FT CONFLICT 328 328 R -> K (IN REF. 3).
 CC FT CONFLICT 426 426 K -> E (IN REF. 3).
 CC FT CONFLICT 429 429 Q -> E (IN REF. 3).
 CC FT CONFLICT 673 675 AQN -> TED (IN REF. 3).
 CC FT CONFLICT 901 901 A -> T (IN REF. 3).
 CC FT CONFLICT 903 903 Q -> E (IN REF. 3).
 CC FT CONFLICT 907 907 L -> P (IN REF. 3).
 CC FT CONFLICT 910 910 P -> R (IN REF. 3).
 CC FT CONFLICT 914 914 Q -> E (IN REF. 3).
 CC FT CONFLICT 1072 1072 P -> S (IN REF. 3).
 CC FT CONFLICT 1074 1074 S -> D (IN REF. 3).
 CC SQ SEQUENCE 1147 AA; 128013 MW; AB92770835f68490 CRC64;
 Query Match 69.6%; Score 32; DB 1; Length 1147;
 Best Local Similarity 83.3%; Pred. No. 91;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NFRYTN 6
 DB 505 NFRYTN 510
 RESULT 9
 ATCT_YEAST STANDARD: PRT; 1151 AA.
 AC P40527;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Potential phospholipid-transporting ATPase 4 (EC 3.6.3.1).
 GN NEO1 OR YIL048W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=S288c / AB972;
 RA Bartell B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,
 RA Church C.M., Connor R., Copley T., Dear S., Devlin K., Fraser A.,
 RA Gentles S., Hamlyn N., Hornell T.S., Hunt S., Jagers K., Jones M.,
 RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,

RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,
 RA Walsh S.V., Whitehead S.; to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
 CC OF ATP COUPLED WITH THE TRANSPORT OF PHOSPHOLIPIDS (POTENTIAL).
 CC LEADS TO NEOMYCIN-RESISTANCE WHEN OVEREXPRESSED.
 CC -1- CATALYTIC ACTIVITY: ATP + H2O -> ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES). SUBFAMILY IV.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z38060; CAA86174.1; -
 CC PIR: S48431; S48431.
 CC SGD: S0001310; NEO1.
 CC InterPro: IPR001757; ATPase_E1-E2.
 CC InterPro: IPR001454; Hlgase/hydrlase.
 CC Pfam: PF00122; E1-E2_ATPase; 1.
 CC Pfam: PF00702; Hydrolase; 1.
 CC DR PROSITE: PS00154; ATPASE_E1-E2; 1.
 CC KW Hydrolyase; Transmembrane; Phosphorylation; Magnesium; ATP-binding.
 CC FT TRANSMEM 185 205 POTENTIAL.
 CC FT TRANSMEM 210 230 POTENTIAL.
 CC FT TRANSMEM 368 388 POTENTIAL.
 CC FT TRANSMEM 417 437 POTENTIAL.
 CC FT TRANSMEM 439 459 POTENTIAL.
 CC FT TRANSMEM 468 498 POTENTIAL.
 CC FT TRANSMEM 971 991 POTENTIAL.
 CC FT TRANSMEM 1021 1041 POTENTIAL.
 CC FT TRANSMEM 1053 1073 POTENTIAL.
 CC FT TRANSMEM 1079 1099 POTENTIAL.
 CC FT TRANSMEM 1110 1130 POTENTIAL.
 CC FT MOD_RES 503 503 POTENTIAL.
 CC SQ SEQUENCE 1151 AA; 130217 MW; DC7225CC9577DB6 CRC64;
 Query Match 69.6%; Score 32; DB 1; Length 1151;
 Best Local Similarity 62.5%; Pred. No. 92;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 NFRYTNFA 8
 DB 106 NFRYTNFA 113
 RESULT 10
 CAGA_HELPJ STANDARD: PRT; 1167 AA.
 ID CAGA_HELPJ
 AC Q9ZLT1;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cytotoxicity associated immunodominant antigen (120 kDa protein)
 DE (CAG pathogenicity island protein 26).
 GN CAGA OR CAI OR CAG26 OR JHP0495.
 GN Helicobacter pylori J99 (Campylobacter pylori J99).
 OS Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=85963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge M.L., Carmel G.,
 RA Tummino P.J., Caruso A., Urita-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;

```

RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE001483; AAD06073.1; -.
DR InterPro: IPR005169; Caga.
DR InterPro: IPR004355; IVSec_caga.
DR Pfam: PF03507; Caga_1.
DR PRINTS: PRO1553; TYPE4SSCAGA.
KW Antigen; Complete proteome.
FT DOMAIN 246 249 POLY-THR.
FT DOMAIN 882 889 POLY-ASN.
SQ SEQUENCE 1167 AA; 129729 MW; FD5E86B81CEBD0F2 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1167;
Best Local Similarity 83.3%; Pred. No. 93;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
Db 509 NFRYTN 514

RESULT 11
CGA2_HELPY STANDARD; PRT; 1182 AA.
ID CGA2_HELPY
AC P5746;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN CAGA OR CAI.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 53726 / 84-183;
RA MEDLINE=93239281; PubMed=8478069;
RA Tumuru M.K.R., Cover T.L., Blaser M.J.;
RT "Cloning and expression of a high-molecular-mass major antigen of
RT Helicobacter pylori: evidence of linkage to cytotoxin production.";
RL Infect. Immun. 61:1799-1809(1993).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: I11714; -. NOT_ANNOTATED_CDS.
DR InterPro: IPR005169; Caga.
DR InterPro: IPR004355; IVSec_caga.
DR Pfam: PF03507; Caga_1.
DR PRINTS: PRO1553; TYPE4SSCAGA.
KW Antigen.
FT DOMAIN 878 885 POLY-ASN.
FT DOMAIN 1182 AA; 131503 MW; C916817E2EB57B4 CRC64;
SQ SEQUENCE 1182 AA; 131503 MW; C916817E2EB57B4 CRC64;

```

```

Query Match 69.6%; Score 32; DB 1; Length 1182;
Best Local Similarity 83.3%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
Db 505 NFRYTN 510

RESULT 12
CGA2_HELPY STANDARD; PRT; 1186 AA.
ID CGA2_HELPY
AC P55980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein)
DE (CAG pathogenically island protein 26).
GN CAGA OR CAI OR CAG26 OR HP0547.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RC MEDLINE=97394467; PubMed=9252185;
RX Tomb J.-F., White O., Kervatage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -!- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
CC OR FUNCTION OF THE CYTOTOXIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000569; AAD07614.1; -.
DR TIGR: HP0547; -.
DR InterPro: IPR005169; Caga.
DR InterPro: IPR004355; IVSec_caga.
DR Pfam: PF03507; Caga_1.
DR PRINTS: PRO1553; TYPE4SSCAGA.
KW Antigen; Complete proteome.
FT DOMAIN 247 250 POLY-THR.
FT DOMAIN 883 889 POLY-ASN.
FT SEQUENCE 1186 AA; 132386 MW; B05C3F2CCC4444F4 CRC64;

Query Match 69.6%; Score 32; DB 1; Length 1186;
Best Local Similarity 83.3%; Pred. No. 95;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTN 6
Db 510 NFRYTN 515

RESULT 13
PGL1_HUMAN

```

```

ID PG11_HUMAN STANDARD; PRT; 377 AA.
AC P53609;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE geranylgeranyl transferase type I beta subunit (EC 2.5.1.-) (Type I
DE protein geranyl-geranyltransferase beta subunit) (GGTase-I-beta).
GN PG11B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta, and Kidney;
RX MEDLINE-94148804; PubMed-8106351;
RA Zhang F.L., Diehl R.E., Kohl N.E., Glibbs J.B., Giros B.,
RA Casey P.J., Omer C.A.;
RT "cDNA cloning and expression of rat and human protein
RT geranylgeranyltransferase type-I."
RL J. Biol. Chem. 269:3175-3180(1994).
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A GERANYL-GERANYL MOIETY FROM
CC GERANYLGERANYL PYROPHOSPHATE TO PROTEINS HAVING THE C-TERMINAL
CC SEQUENCE CYS-ALIPHATIC-ALIPHATIC-X. ACTS ON THE RAC1, RAC2,
CC RAP1A AND RAP1B PROTEINS.
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 4 PFTB REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: L25441; AAA5888.1; -
DR Genew: HGNC:8895; PGCT1B.
DR MIM: 602031;
DR InterPro: IPR001330; Prenyltrans.
DR Pfam: PF00432; Prenyltrans. 5.
KW Transferase; Prenyltransferase; Repeat; Zinc.
FT REPEAT 144 186 PFTB 1.
FT REPEAT 193 234 PFTB 2.
FT REPEAT 245 284 PFTB 3.
FT REPEAT 291 333 PFTB 4.
FT METAL 269 269 ZINC (BY SIMILARITY).
FT METAL 271 271 ZINC (BY SIMILARITY).
FT METAL 321 321 ZINC (BY SIMILARITY).
SQ SEQUENCE 377 AA; 42396 MW; 565CD9B6C087AADF CRC64;

Query Match 67.4%; Score 31; DB 1; Length 377;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 FRYTNF 7
Db 286 FORTNF 291

RESULT 14
PG11_BAT STANDARD; PRT; 377 AA.
AC P53610;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Geranylgeranyl transferase type I beta subunit (EC 2.5.1.-) (Type I
DE protein geranyl-geranyltransferase beta subunit) (GGTase-I-beta).
GN PG11B.

```

```

OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE-94148804; PubMed-8106351;
RA Zhang F.L., Diehl R.E., Kohl N.E., Glibbs J.B., Giros B.,
RA Casey P.J., Omer C.A.;
RT "cDNA cloning and expression of rat and human protein
RT geranylgeranyltransferase type-I."
RL J. Biol. Chem. 269:3175-3180(1994).
CC -1- FUNCTION: CATALYZES THE TRANSFER OF A GERANYL-GERANYL MOIETY FROM
CC GERANYLGERANYL PYROPHOSPHATE TO PROTEINS HAVING THE C-TERMINAL
CC SEQUENCE CYS-ALIPHATIC-ALIPHATIC-X. ACTS ON THE RAC1, RAC2,
CC RAP1A AND RAP1B PROTEINS.
CC -1- COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA SUBUNIT.
CC -1- SIMILARITY: BELONGS TO THE PROTEIN PRENYLTRANSFERASE BETA SUBUNIT
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 4 PFTB REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL: L24116; AAA17756.1; -
DR InterPro: IPR001330; Prenyltrans.
DR Pfam: PF00432; Prenyltrans. 5.
KW Transferase; Prenyltransferase; Repeat; Zinc.
FT REPEAT 144 186 PFTB 1.
FT REPEAT 193 234 PFTB 2.
FT REPEAT 245 284 PFTB 3.
FT REPEAT 291 333 PFTB 4.
FT METAL 269 269 ZINC (BY SIMILARITY).
FT METAL 271 271 ZINC (BY SIMILARITY).
FT METAL 321 321 ZINC (BY SIMILARITY).
SQ SEQUENCE 377 AA; 42414 MW; AADEC7301A4A4011 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 377;
Best Local Similarity 83.3%; Pred. No. 47;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 FRYTNF 7
Db 286 FORTNF 291

RESULT 15
CKAA_BACUF STANDARD; PRT; 753 AA.
AC Q32321;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pesticidal crystal protein cry20Aa (insecticidal delta-endotoxin
DE cryXXAa) (Crystalline entomocidal protoxin) (66 kDa crystal
DE protein).
GN CRY20AA OR CRYXXA(A).
OC Bacillus thuringiensis (subsp. fukuokaensis).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=132265;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-98069459; PubMed-9406385;
RA Lee H.-K., Gill S.S.;
RT "Molecular cloning and characterization of a novel mosquitocidal
RT protein gene from Bacillus thuringiensis subsp. fukuokaensis.";

```

RL Appl. Environ. Microbiol. 63:4664-4670(1997).
CC -1- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
CC EPITHELIAL CELLS OF MOSQUITOS. ACTIVE AGAINST AEDES AEGYPTI AND
CC CULEX QUINQUEFASCIATUS LARVAE.
CC -1- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
CC SPORELIATION AND IS ACCUMULATED BOTH AS AN INCLUSION AND AS PART
CC OF THE SPORE COAT.
CC -1- PTM: HAS LOW MOSQUITOCIDAL ACTIVITY PROBABLY DUE TO RAPID
CC PROTEOLYSIS TO INACTIVE 56-KDA AND 43-KDA PROTEINS.
CC -1- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
CC N-TERMINUS.
CC -1- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U82518; AAB93476.1; -
DR InterPro: IPR001178; Endotoxin.
DR Pfam: PF00555; endotoxin; 1.
DR Toxin; Sporulation.
KW SEQUENCE 753 AA; 86139 MW; CE3DED573EA41D66 CRC64;

Query Match 67.4%; Score 31; DB 1; Length 753;
Best Local Similarity 62.5%; Pred. No. 94;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 NFRYTNEA 8
||:| ||
Db 591 NFRYTNEA 598

Search completed: March 25, 2003, 08:20:14
Job time : 7.45455 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:19:47 ; Search time 19.9091 Seconds
(without alignments)
93.145 Million cell updates/sec

Title: US-09-646-532b-3

Perfect score: 46

Sequence: 1 NFRYTNEFX 9

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirts:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Match Length | ID | Description |
|------------|-------|-------|--------------|----|-------------|
| 1 | 36 | 78.3 | 453 | 10 | Q9LTI7 |
| 2 | 36 | 78.3 | 457 | 10 | Q93Y40 |
| 3 | 34 | 73.9 | 222 | 16 | Q9KOL7 |
| 4 | 34 | 73.9 | 2831 | 16 | Q8UBX0 |
| 5 | 34 | 73.9 | 2832 | 2 | Q93N05 |
| 6 | 33 | 71.7 | 99 | 1 | Q9UXD1 |
| 7 | 33 | 71.7 | 104 | 16 | Q8UB48 |
| 8 | 33 | 71.7 | 157 | 12 | Q98434 |
| 9 | 33 | 71.7 | 262 | 5 | Q9XU29 |
| 10 | 33 | 71.7 | 432 | 12 | Q90311 |
| 11 | 33 | 71.7 | 436 | 12 | Q9WNU3 |
| 12 | 33 | 71.7 | 437 | 12 | Q36189 |
| 13 | 33 | 71.7 | 437 | 12 | Q9Y228 |
| 14 | 33 | 71.7 | 476 | 12 | Q80MG3 |
| 15 | 33 | 71.7 | 511 | 16 | Q51113 |
| 16 | 33 | 71.7 | 520 | 12 | Q41104 |

| | | | | | | |
|----|----|------|------|----|--------|--------------------|
| 17 | 33 | 71.7 | 521 | 12 | Q90312 | 090312 chlorella v |
| 18 | 33 | 71.7 | 950 | 5 | Q96360 | 096360 heliothis v |
| 19 | 33 | 71.7 | 952 | 5 | Q9NH28 | Q9NH28 heliothis v |
| 20 | 33 | 71.7 | 1498 | 4 | Q9P267 | Q9P267 homo sapien |
| 21 | 33 | 71.7 | 1960 | 5 | Q900K8 | Q900K8 plasmidium |
| 22 | 33 | 71.7 | 2115 | 17 | Q870N0 | Q870N0 methanosarc |
| 23 | 32 | 69.6 | 83 | 12 | Q8V588 | Q8V588 tomato yell |
| 24 | 32 | 69.6 | 83 | 12 | Q8V584 | Q8V584 tomato yell |
| 25 | 32 | 69.6 | 83 | 12 | Q8V582 | Q8V582 tomato yell |
| 26 | 32 | 69.6 | 83 | 12 | Q8V580 | Q8V580 tomato yell |
| 27 | 32 | 69.6 | 83 | 12 | Q8V578 | Q8V578 tomato yell |
| 28 | 32 | 69.6 | 83 | 12 | Q8V576 | Q8V576 tomato yell |
| 29 | 32 | 69.6 | 115 | 12 | Q8V612 | Q8V612 tomato yell |
| 30 | 32 | 69.6 | 209 | 12 | Q8VAC5 | Q8VAC5 white spot |
| 31 | 32 | 69.6 | 258 | 5 | Q45856 | Q45856 caenorhabdi |
| 32 | 32 | 69.6 | 270 | 11 | Q8RIE1 | Q8RIE1 mus musculu |
| 33 | 32 | 69.6 | 346 | 16 | Q98PU2 | Q98PU2 mycoplasma |
| 34 | 32 | 69.6 | 369 | 10 | Q9SHH4 | Q9SHH4 arabidopsis |
| 35 | 32 | 69.6 | 370 | 11 | Q91YX5 | Q91YX5 mus musculu |
| 36 | 32 | 69.6 | 467 | 10 | P93709 | P93709 nicotiana t |
| 37 | 32 | 69.6 | 563 | 17 | Q8TP23 | Q8TP23 methanosarc |
| 38 | 32 | 69.6 | 587 | 5 | Q8T2T9 | Q8T2T9 dictyostell |
| 39 | 32 | 69.6 | 595 | 10 | Q9SMQ2 | Q9SMQ2 arabidopsis |
| 40 | 32 | 69.6 | 599 | 10 | Q94BN1 | Q94BN1 arabidopsis |
| 41 | 32 | 69.6 | 604 | 17 | Q51989 | Q51989 halobacteri |
| 42 | 32 | 69.6 | 604 | 17 | Q51959 | Q51959 halobacteri |
| 43 | 32 | 69.6 | 607 | 16 | Q9RX8 | Q9RX8 delnoccocu |
| 44 | 32 | 69.6 | 713 | 5 | P91982 | P91982 caenorhabdi |
| 45 | 32 | 69.6 | 912 | 11 | Q70410 | Q70410 mus musculu |

ALIGNMENTS

RESULT 1
ID Q9LTI7 PRELIMINARY; PRT: 453 AA.
AC Q9LTI7
DT 01-OCT-2000 (TREMUREL. 15, Created)
DT 01-OCT-2000 (TREMUREL. 15, Last sequence update)
DT 01-JUN-2001 (TREMUREL. 17, Last annotation update)
DE Oxytelol-binding protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneo T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025604; BAA97478.1; -;
DR InterPro: IPR000648; Oxytelol_BP.
DR Pfam: PF01237; Oxytelol_BP.1.
DR PROSITE: PS01013; OSBP.1.
SQ SEQUENCE 453 AA; 51507 MW; FB326FF0088D1EC CRC64;

Query Match 78.3%; Score 36; DB 10; Length 453;
Best Local Similarity 75.0%; Pred. NO. 38;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 NFRYTNEFX 8
DB 323 NFRYTNEFX 330

RESULT 2
ID Q93Y40 PRELIMINARY; PRT: 457 AA.
AC Q93Y40;

DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Oxyesterol-binding protein.
 GN AT5G59350 OR P2015.10.
 OC Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carlinici P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamliya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shim P., Yamada K., Shinozaki K.,
 RA Becker J., Theologis A., Davis R.W.;
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY054473; AAK96664.1; -
 DR InterPro: IPR000648; Oxysterol_BP.
 DR Pfam: PF01237; Oxysterol_BP.
 DR PROSITE: PS01013; OSRP: UNKNOWN.1.
 SQ SEQUENCE 457 AA; 51966 MW; 90DDB0654110D85D CRC64;

Query Match 78.3%; Score 36; DB 10; Length 457;
 Best Local Similarity 75.0%; Pred. No. 38;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNA 8
 ||:||||
 DB 327 NFOYTHFA 334

RESULT 3
 ID 09KOL7 PRELIMINARY; PRT; 222 AA.
 AC 09KOL7;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
 DE Hypothetical protein VC1981.
 GN VC1981.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OC NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=EL TOR N16961 / SEROTYPE O1;
 RC MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
 RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 RA Fraser C.M.;
 RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae."
 RL Nature 406:477-483(2000).
 DR EMBL: AE004273; AAF95129.1; -
 DR TIGR: VC1981; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 222 AA; 24813 MW; 4FBC63ED9E5214F2 CRC64;

Query Match 73.9%; Score 34; DB 16; Length 222;
 Best Local Similarity 75.0%; Pred. No. 46;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 NFRYTNA 8
 ||:||||
 DB 84 NFAHTNFA 91

RESULT 4
 ID 08UBX0 PRELIMINARY; PRT; 2831 AA.
 AC 08UBX0;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE Beta (1-->2) glucan biosynthesis protein.
 GN CHVB OR ATU2730 OR AGR_C_4949.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bevee D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gilliet W., Grant C.,
 RA Kuttyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Lao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 RT C58."
 RL Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmel K., Gordon J., Vaudin M., Tarchoux O., Epp A., Liu F.,
 RA Woliam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursun J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RL Science 294:2323-2328(2001).
 DR EMBL: AE009220; AAL43711.1; -
 DR EMBL: AE008186; AAK84447.1; ALT_INIT.
 KW Complete proteome.
 SQ SEQUENCE 2831 AA; 316342 MW; 08E9C1F99A763FC6 CRC64;

Query Match 73.9%; Score 34; DB 16; Length 2831;
 Best Local Similarity 71.4%; Pred. No. 5; 7e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNA 7
 ||:||||
 DB 1397 NYQYTNF 1403

RESULT 5
 ID 093N05 PRELIMINARY; PRT; 2832 AA.
 AC 093N05;
 DT 01-DEC-2001 (Tremblrel. 19, Created)
 DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE Cyclic beta 1-2 glucan synthetase.
 GN CHVB.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 OC NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C58;
 RA Iannino F., Inon de Iannino N.;

RT "Cyclic glucan synthetase of Agrobacterium tumefaciens, comparative
sequence analysis." ;
RL Submitted (JUN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AK7391087; AAK73356.1; -
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR000130; Zn_MTPetase.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
SQ SEQUENCE 2832 AA; 316536 MW; BA52584BA3C1ABA CRC64;

Query Match 73.9%; Score 34; DB 2; Length 2832;
Best Local Similarity 71.4%; Pred. No. 5.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
Db 1397 NQOYTNE 1403

RESULT 6
O9UXD1 PRELIMINARY; PRT; 99 AA.
AC O9UXD1:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE Hypothetical protein 11.1 kDa protein.
OS ORF-C20.035.
NS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1617 / P2.
RA Charlebois R.L., Singh R.K., Chan-Welher C.C.-Y., Allard G., Chow C.,
RA Contalonieli F., Curtis B., Duguet M., Erauso G., Faguy D.,
RA Gaasterland T., Garrett R.A., Gordon P., Jeffries A.C., Kozera C.,
RA Kushwaha N., Lafleur E., Medina N., Peng X., Penny S.L., She Q.,
RA St Jean A., Van Der Oost J., Young F., Zivanovic Y., Doolittle W.F.,
RA Ragan M.A., Sensen C.W.;
RT "Gene content and organization of a 281-kbp contig from the genome of
the extremely thermophilic archaeon, Sulfolobus solfataricus P2." ;
RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: Y18930; CAB57560.1; -
KW Hypothetical protein.
SQ SEQUENCE 99 AA; 11090 MW; 83341F51627C1AB4 CRC64;

Query Match 71.7%; Score 33; DB 1; Length 99;
Best Local Similarity 71.4%; Pred. No. 33;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
Db 47 NFRFSNF 53

RESULT 7
O8UE48 PRELIMINARY; PRT; 104 AA.
AC O8UE48:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Hypothetical protein Atul1913.
GN Atul1913 OR AGC_C_3504.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RX

RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dotan M.,
RA Chunley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58." ;
RT Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Guroillo B., Goldman B.S., Cao Y., Askenazi M., Hailing C., Mullin L.,
RA Houtiel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Guron J., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58." ;
RL Science 294:2323-2328(2001).
DR EMBL: AE009145; AAI42909.1; -
DR EMBL: AE008110; AAK87673.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 104 AA; 11642 MW; 8C5505A58E69A308 CRC64;

Query Match 71.7%; Score 33; DB 16; Length 104;
Best Local Similarity 71.4%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 NFRYTNF 7
Db 5 NFRYTHY 11

RESULT 8
O98434 PRELIMINARY; PRT; 157 AA.
AC O98434:
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE A382R protein.
GN A382R.
OS Paramesidium bursaria chlorella virus 1 (PBCV-1).
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10506;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96400190; PubMed=8806566;
RA Kutish G.F., Li Y., Lu Z., Furuta M., Rock D.L., Van Etten J.L.;
RT "Analysis of 76 kb of the chlorella virus PBCV-1 330-kb genome: map
RT positions 182 to 258." ;
RL Virology 223:303-317(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20013326; PubMed=10544099;
RA Kaiser A.D., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
RA Lisek A.D., Nickerson K.W., Van Etten J.L.;
RT "Chlorella virus PBCV-1 encodes a functional homospesmidine
RT synthase." ;
RL Virology 263:254-262(1999).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20478054; PubMed=11021991;
RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
RT PBCV-1." ;
RL Virology 276:27-36(2000).
RX

[4]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Van Etten J.L.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RA Graves M.V., Van Etten J.L.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [9]
RP SEQUENCE FROM N.A.
RA Gunton J.R., Graves M.V., Van Etten J.L.;
RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42580; AAC96750.1; -;
SQ SEQUENCE 157 AA; 18046 MW; 25EF8B6122E1014E CRC64;

Query Match
Best Local Similarity 71.7%; Score 33; DB 12; Length 157;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RYTNEA 8
Db 125 RYTNEA 130

RESULT 9
Q9XU29 PRELIMINARY; PRT; 262 AA.
ID Q9XU29
AC Q9XU29;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE F47H4.8 protein.
GN F47H4.8.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Stewart C.A.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL; Z92786; CAB07207.1; -;
DR InterPro; IPR002900; DUF38.
DR InterPro; IPR001810; F-box.
DR Pfam; PF01827; DUF38.1.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR PROSITE; PS50181; FBOX; 1.
SQ SEQUENCE 262 AA; 13280 MW; B8F5B2065558EC8 CRC64;

Query Match
Best Local Similarity 71.7%; Score 33; DB 5; Length 262;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 NERYTNE 7
Db 206 NERYTNE 212

RESULT 10
ID Q90311 PRELIMINARY; PRT; 432 AA.
AC Q90311;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Major capsid protein Vp49.
GN VCR7.
OS Chlorella virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CVG-1;
RX MEDLINE=99226959; PubMed=10211977;
RA Plugge B., Becker B., Wolf A.H.;
RT "Several genes in Chlorella virus strain CVG-1 encode putative virion
components."
RL J. Gen. Virol. 80:1067-1072(1999).
DR EMBL; AF076921; AAC27492.1; -;
DR InterPro; IPR002106; ATRNA_ligaseII.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II.1; UNKNOWN.1.
SQ SEQUENCE 432 AA; 47696 MW; FD9732A4F7876D1 CRC64;

Query Match
Best Local Similarity 71.7%; Score 33; DB 12; Length 432;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RYTNEA 8
Db 33 RYTNEA 38

RESULT 11
ID Q9WNU3 PRELIMINARY; PRT; 436 AA.
AC Q9WNU3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE Major capsid protein.
GN CVK2MCP-2.
OS Chlorella virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CVK2;
RX MEDLINE=99170731; PubMed=10069963;
RA Nishida K., Kimura Y., Kawasaki T., Fujie M., Yamada T.;
RT "Genetic variation of Chlorella viruses: Variable regions localized on
the CVK2 genomic DNA."
RL Virology 255:376-384(1999).
DR EMBL; AB018568; BAA76600.1; -;
SQ SEQUENCE 436 AA; 48398 MW; 01DAE208040C8D82 CRC64;

Query Match
Best Local Similarity 71.7%; Score 33; DB 12; Length 436;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RYTNEA 8
Db 33 RYTNEA 38

RESULT 12
Q36189

```

ID 036189 PRELIMINARY: PRT: 437 AA.
AC 036189;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)
DE Major capsid protein Vp54.
GN Vp54.
OS Chlorella virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=10507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CVT2;
RX MEDLINE=98190351; PubMed=9514979;
RA Nishida K., Suzuki S., Kimura Y., Nomura N., Fujie M., Yamada T.;
RT "Group I introns found in Chlorella viruses: Biological
RT implications."
RL Virology 242:319-326(1998).
DR EMBL; AB006978; BAA22198.1; -.
SQ SEQUENCE 437 AA; 48049 MW; 53FEA9AE971F3683 CRC64;

```

```

Query Match 71.7%; Score 33; DB 12; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 RYTNEA 8
DB 33 RYTNEA 38

```

```

RESULT 13
O9YZZ8 PRELIMINARY: PRT: 437 AA.
ID 09YZZ8;
AC 09YZZ8;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE Major capsid protein VP52.
GN VP52.
OS Chlorella virus K2, and
OS Chlorella virus K2.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Chlorovirus.
OX NCBI_TaxID=75912, 31555;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Chlorella virus K2;
RA Yamada T., Nishida K., Fujie M., Usami S.;
RT "Algal viruses use their own tRNAs expressed in host cells for the
RT protein synthesis."
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=Chlorella virus CVK2;
RX MEDLINE=99170731; PubMed=10069963;
RA Nishida K., Kimura Y., Kawasaki T., Fujie M., Yamada T.;
RT "Genetic variation of Chlorella viruses: variable regions localized on
RT the CVK2 genomic DNA."
RL Virology 255:376-384(1999).
DR EMBL; AB011506; BAA35143.1; -.
DR EMBL; AB018579; BAA76601.1; -.
SQ SEQUENCE 437 AA; 48280 MW; ACA83D9F235AF2D CRC64;

```

```

Query Match 71.7%; Score 33; DB 12; Length 437;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 RYTNEA 8
DB 33 RYTNEA 38

```

```

RESULT 14
O8ONG3

```

```

ID 08ONG3 PRELIMINARY: PRT: 476 AA.
AC 08ONG3;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Esv-1-116.
GN Esv-1-116.
OS Ectocarpus siliculosus virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phaeovirus.
OX NCBI_TaxID=37665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ESV-1;
RA Delaroque N., Bothe G., Pohl T., Knippers R., Mueller D.G., Boland W.;
RT "The complete nucleotide sequence of the Ectocarpus siliculosus virus
RT genome."
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF204951; AAK14534.1; -.
SQ SEQUENCE 476 AA; 53212 MW; A6F5C92F3C92526D CRC64;

```

```

Query Match 71.7%; Score 33; DB 12; Length 476;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 3 RYTNEA 8
DB 10 RYTNEA 15

```

```

RESULT 15
O51113 PRELIMINARY: PRT: 511 AA.
ID 051113;
AC 051113;
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Hypothetical protein BB0086.
GN BB0086.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kellavag A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Uitterback T., Matthews L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
DR EMBL; AE001121; AAC66471.1; -.
DR TIGR; BB0086; -.
DR InterPro; IPR000523; Mg_chelatase-ChlI.
DR InterPro; IPR004482; Mg_chelatase; 1.
DR Pfam; PF01078; Mg_chelatase; 1.
DR TIGRFAMs; TIGR00368; Mg_chelat_rel; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 511 AA; 57237 MW; 18E7514CBE6D0B5 CRC64;

```

```

Query Match 71.7%; Score 33; DB 16; Length 511;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 NRYTNEA 8
DB 423 NRYTNEA 430

```

Fri Mar 28 10:51:43 2003

us-09-646-532b-3.rspt

Page 6

Search completed: March 25, 2003, 08:21:37
Job time : 21.9091 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:17:33 ; Search time 40.7273 Seconds
(without alignments)
45.805 Million cell updates/sec

Title: US-09-646-532B-2
Perfect score: 62
Sequence: 1 GAXXXETAMACGXA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------|--------------------|
| 1 | 54 | 87.1 | 14 | AAV42754 | Wheat amyloplast A |
| 2 | 54 | 87.1 | 14 | AAV39332 | ADP glucose transp |
| 3 | 50 | 80.6 | 12 | AAW95412 | Integrin-specific |
| 4 | 50 | 80.6 | 12 | AAE17101 | Cyclic Integrin-bl |
| 5 | 50 | 80.6 | 13 | AAE17101 | Integrin binding p |
| 6 | 50 | 80.6 | 13 | AAE17101 | Alpha-5-beta-1 int |
| 7 | 50 | 80.6 | 23 | AAU74986 | Transfection assoc |
| 8 | 50 | 80.6 | 23 | AAE17116 | Cyclic Integrin-bl |
| 9 | 50 | 80.6 | 25 | AAW95416 | Integrin-binding p |
| 10 | 50 | 80.6 | 25 | AAU74968 | Non-RGD containing |

| | | | | | |
|----|----|------|-----|-----------|--------------------|
| 11 | 50 | 80.6 | 25 | AAE17106 | Integrin-binding p |
| 12 | 50 | 80.6 | 28 | AAU74971 | Alpha 5 beta 1 int |
| 13 | 50 | 80.6 | 28 | AAE17121 | Integrin-binding o |
| 14 | 50 | 80.6 | 31 | AAU74976 | Integrin binding o |
| 15 | 44 | 71.0 | 10 | AAU74982 | Transfection assoc |
| 16 | 44 | 71.0 | 10 | AAE17112 | Cyclic Integrin-bl |
| 17 | 44 | 71.0 | 12 | AAW95411 | Integrin-specific |
| 18 | 44 | 71.0 | 12 | AAU74961 | Non-RGD containing |
| 19 | 44 | 71.0 | 12 | AAE17100 | Cyclic Integrin-bl |
| 20 | 44 | 71.0 | 30 | AAE17118 | Integrin-binding o |
| 21 | 40 | 64.5 | 769 | AAE17118 | Novel human diagno |
| 22 | 39 | 62.3 | 669 | AAE10335 | Human transporter |
| 23 | 38 | 61.3 | 9 | AAE17106 | Alpha5/beta1 integ |
| 24 | 38 | 61.3 | 9 | AAW57200 | RGD-containing pep |
| 25 | 38 | 61.3 | 9 | AAW95413 | Integrin-binding p |
| 26 | 38 | 61.3 | 9 | AAV56589 | Integrin alphasbet |
| 27 | 38 | 61.3 | 9 | AAE17986 | Human ligand #6 at |
| 28 | 38 | 61.3 | 9 | AAU74963 | Non-RGD containing |
| 29 | 38 | 61.3 | 10 | AAU74980 | Transfection assoc |
| 30 | 38 | 61.3 | 10 | AAE17111 | Cyclic Integrin-bl |
| 31 | 38 | 61.3 | 13 | AAW56041 | Chimeric adenoviru |
| 32 | 37 | 59.7 | 38 | ABB37805 | Peptide #5311 enco |
| 33 | 37 | 59.7 | 38 | ABB23080 | Protein #5079 enco |
| 34 | 37 | 59.7 | 38 | AAW58421 | Human brain expres |
| 35 | 37 | 59.7 | 38 | AAW70909 | Human bone marrow |
| 36 | 37 | 59.7 | 38 | AAW18729 | Peptide #5163 enco |
| 37 | 37 | 59.7 | 38 | AAW31201 | Peptide #5238 enco |
| 38 | 37 | 59.7 | 38 | ABG40705 | Human peptide enco |
| 39 | 37 | 59.7 | 38 | ABG60196 | Human DTPAP poly |
| 40 | 37 | 59.7 | 433 | ABG16929 | Novel human diagno |
| 41 | 36 | 58.1 | 76 | ABP04697 | Human ORFX protein |
| 42 | 36 | 58.1 | 116 | ABG02355 | Human secreted pro |
| 43 | 36 | 58.1 | 129 | AAE12569 | zee mays protein f |
| 44 | 36 | 58.1 | 157 | AAE08461 | Novel human diagno |
| 45 | 36 | 58.1 | 157 | AAE072946 | Human olfactory re |

ALIGNMENTS

| | | |
|----------|--|-------------------------------|
| RESULT 1 | AAV42754 | standard; peptide; 14 AA. |
| ID | AAV42754 | standard; peptide; 14 AA. |
| XX | AAV42754 | |
| AC | AAV42754 | |
| XX | | |
| DT | 20-DEC-1999 | (first entry) |
| XX | | |
| DE | Wheat amyloplast ADP-glucose transporter peptide #2. | |
| XX | | |
| KW | Starch biosynthesis; amyloplast; ADP-glucose; transport; import; | |
| KW | amylopectin; amylose; branching; chemical structure; transgenic plant; | |
| KW | optimisation; industrial applications. | |
| XX | | |
| OS | Triticum aestivum. | |
| XX | | |
| FT | Key | Location/Qualifiers |
| FT | Misc-difference 3 | /label= Xaa |
| FT | Misc-difference 4 | /note= "Xaa = any amino acid" |
| FT | Misc-difference 5 | /label= Xaa |
| FT | Misc-difference 5 | /note= "Xaa = any amino acid" |
| FT | Misc-difference 13 | /label= Xaa |
| FT | Misc-difference 13 | /note= "Xaa = any amino acid" |
| FT | Misc-difference 13 | /label= Xaa |
| FT | Misc-difference 13 | /note= "Xaa = any amino acid" |
| PN | WO9947682-A1. | |
| XX | | |
| PD | 23-SEP-1999. | |

Integrin-binding p
Alpha 5 beta 1 int
Integrin-binding o
Integrin binding o
Transfection assoc
Cyclic Integrin-bl
Integrin-specific
Non-RGD containing
Cyclic Integrin-bl
Integrin-binding o
Novel human diagno
Human transporter
Alpha5/beta1 integ
RGD-containing pep
Integrin-binding p
Integrin alphasbet
Human ligand #6 at
Non-RGD containing
Transfection assoc
Cyclic Integrin-bl
Chimeric adenoviru
Peptide #5311 enco
Protein #5079 enco
Human brain expres
Human bone marrow
Peptide #5163 enco
Peptide #5238 enco
Human peptide enco
Human DTPAP poly
Novel human diagno
Human ORFX protein
Human secreted pro
zee mays protein f
Novel human diagno
Human olfactory re

```

XX 19-MAR-1999; 99WO-GB00728.
PF 20-MAR-1998; 98GB-0005939.
XX (UYMA-) UNIV VICTORIA MANCHESTER.
XX Emes MJ, Tellow IJ, Bowersher CG;
XX WPI: 1999-590977/50.
DR New transporter complex protein useful for modulating starch content in
XX plants, especially useful in food production.
XX Claim 1; Page 3; 28pp; English.
XX This sequence represents a wheat amyloplast ADP-glucose transporter
CC peptide, #2. The wheat amyloplast ADP-glucose transporter is associated
CC with the amyloplast membrane and comprises at least two proteins; this
CC sequence, along with peptides #1 (AAV42753) and #3 (AAV42755) are
CC components of one of these proteins. Peptides #4-#7 (AAV42756-Y42759)
CC represent sequences within the second protein (AAV42760). The sugar
CC nucleotide ADP-glucose is the immediate substrate for starch synthesis,
CC which occurs in the amyloplast; however, ADP-glucose is mainly
CC synthesised outside the amyloplast in the cytoplasm. The ADP-glucose
CC transporter is responsible for the import of ADP-glucose into the
CC amyloplast and therefore plays a pivotal role in the regulation of starch
CC synthesis. The transporter not only influences starch yield, but also
CC quality as the starch synthases involved in amylose and amylopectin
CC synthesis have different affinities for ADP-glucose. Variations in the
CC chemical structure of starch are determined by the ratio of amylose to
CC amylopectin, and by the degree of branching in amylopectin in the starch
CC polymer. These variations can significantly alter the properties of
CC starch. The ADP-glucose transporter complex is useful for generating
CC transgenic plants in which the starch quality can be optimised for
CC industrial applications in which starch is used. For example, transgenic
CC plants which have an increased amylose content in starch are useful for
CC production of starch with increased viscosity and gel strength, which
CC prevents baked food going stale so quickly. Conversely, an increased
CC amylopectin content in the starch produces waxy starch useful as
CC thickening agents in food and coatings.
XX Sequence 14 AA:
SQ
Query Match 87.1%; Score 54; DB 20; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAXXXETAMACGXA 14
Db 1 GAXXXETAMACGXA 14
RESULT 2
ID AAV39332 standard; peptide: 14 AA.
XX AAV39332;
AC 01-DEC-1999 (first entry)
XX ADP glucose transporter peptide 2.
DE ADP glucose transporter peptide 2.
XX ADP glucose transporter; transform plant cell; wheat; starch production;
KW waxy starch; thickening agent; food; coating; increased viscosity; state;
KW gel strength; baked food.
XX Triticum aestivum.
OS
XX Key Location/Qualifiers
FH Misc-difference 3
FT /note= "Any amino acid"
FT Misc-difference 4

```

```

FT /note= "Any amino acid"
FT Misc-difference 5
FT /note= "Any amino acid"
FT Misc-difference 13
FT /note= "Any amino acid"
XX WO9947681-A1.
XX 23-SEP-1999.
XX 19-MAR-1999; 99WO-GB00727.
XX 20-MAR-1998; 98GB-0005939.
XX (UYMA-) UNIV VICTORIA MANCHESTER.
XX Emes MJ, Tellow IJ, Bowersher CG;
XX WPI: 1999-571841/48.
XX ADP glucose transporter protein used for modifying plant starch
XX production.
XX Claim 1; Page 15; 26pp; English.
XX Peptides AAV39331-Y39337 are fragments of an ADP glucose transporter
CC and protein. The protein contains at least one of the peptide sequences, and
CC is capable of ADP glucose transport. A DNA molecule encoding an ADP
CC glucose transporter protein can be used to transform plant cells. The
CC ADP glucose transporter protein can be used to regulate starch
CC production from a plant. The plants can then be used to produce waxy
CC starches that can be used as thickening agents in food and coatings.
CC Alternatively the plants can be used to produce a starch with increased
CC viscosity and gel strength, which can be incorporated in baked food which
CC stays fresh for a longer length of time.
XX Sequence 14 AA:
SQ
Query Match 87.1%; Score 54; DB 20; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.0091;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 GAXXXETAMACGXA 14
Db 1 GAXXXETAMACGXA 14
RESULT 3
ID AAW95412 standard; peptide: 12 AA.
XX AAW95412;
AC 18-MAR-1999 (first entry)
XX Integrin-specific peptide.
DE Integrin-specific peptide.
XX Integrin; transfection complex; integrin-binding; lipid; immunisation;
KW antisense therapy; enzyme; therapeutic agent; immunogen; cystic fibrosis;
KW cancer; viral infection; human immunodeficiency virus; cardiovascular;
KW resenosis; leukaemia; asthma; glaucoma.
XX Synthetic.
OS
XX WO9854347-A1.
XX 03-DEC-1998.
XX 29-MAY-1998; 98WO-GB01577.
XX 29-MAY-1997; 97GB-0011115.
XX (CHIL-) INST CHILD HEALTH.

```


XX Hart SL;
 PI
 XX
 DR MPI: 1999-045366/04.
 XX
 PT New integrin-targeting transfection complex including lipid - used
 PT to improve transfection efficiency for a very wide range of cells,
 PT useful in, e.g. antisense therapy and genetic immunisation
 XX
 PS Claim 11; Page 50; 70pp; English.
 XX
 CC The invention relates to an integrin-targeting transfection complex. The
 CC complex comprises a nucleic acid, an integrin-binding component, a
 CC polycationic nucleic acid-binding component and a lipid. The complexes
 CC are used for in vivo or in vitro transfection of cells, specifically:
 CC (1) for treatment or prevention of disease (in humans or other animals)
 CC caused by defective or deficient genes; (11) for immunisation; (11i) for
 CC antisense therapy; and (iv) for protein production in host cells, e.g.
 CC of enzymes, therapeutic agents, vaccinating immunogens and diagnostic
 CC antigens. Typical of the diseases that can be treated or prevented are
 CC cystic fibrosis, cancer, viral infection (e.g. human immunodeficiency
 CC virus), cardiovascular disease (e.g. restenosis), leukaemia, asthma and
 CC glaucoma. Incorporation of the lipid into the complex increases
 CC transfection levels from 1-10 percent to over 50 percent. This effect is
 CC observed with all cell types tested including those that are resistant to
 CC transfection by most plasmid vectors. The complexes can carry large
 CC genes, up to 125 kb, e.g. an artificial chromosome. The present sequence
 CC represents a claimed example of an integrin-binding peptide used in the
 CC transfection complexes.
 CC
 XX Sequence 12 AA:
 SQ
 Query Match 80.6%; Score 50; DB 20; Length 12;
 Best Local Similarity 75.0%; Pred. No. 0.037;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GAXXXETAWACG 12
 II |
 DB 1 GACRRETAWACG 12
 RESULT 4
 AAEL7101
 ID AAEL7101 standard; peptide: 12 AA.
 XX
 AC AAEL7101;
 XX
 DT 18-APR-2002 (first entry)
 XX
 DE Cyclic integrin-binding peptide 6.
 XX
 KW Integrin binding component; polycationic nucleic acid-binding component;
 KW lipid component; prophylaxis; immunisation; antisense therapy; asthma;
 KW cystic fibrosis; cancer; viral infection; human immunodeficiency virus;
 KW HIV infection; vaccine; neuroblastoma; bone marrow stem cell disorder;
 KW leukaemia; adjuvant immunotherapy; eye disease; glaucoma; restenosis;
 KW integrin-binding peptide; cyclic.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Disulfide-bond 3.11
 FT /note="Forms a cyclic structure"
 XX
 PN WO200192542-A2.
 XX
 PD 06-DEC-2001.
 PD
 PF 30-MAY-2001; 2001WO-CB02394.
 PF
 PR 30-MAY-2000; 2000GB-0013089.
 PR 30-MAY-2000; 2000GB-0013090.
 PR 01-MAY-2001; 2001US-287410P.

XX (ICHI-) ICH PRODN LTD.
 PA
 XX
 PI Hart SL;
 XX
 DR MPI: 2002-139612/18.
 XX
 PT Complex for transfecting cell with nucleic acid for treating,
 PT preventing conditions caused by deficiency in a gene in humans, has
 PT nucleic acid, lipid, integrin binding and polycationic nucleic
 PT acid-binding components
 XX
 PS Claim 17; Page 6; 108pp; English.
 XX
 CC The invention relates to integrin-targeting vectors having enhanced
 CC transfection activity. The vector complex comprises a nucleic acid,
 CC an integrin binding component, a polycationic nucleic acid-binding
 CC component and a lipid component. The integrin binding component
 CC comprises an integrin-binding element and a spacer element. Complex
 CC of the invention is useful for transfecting cells in vitro or in
 CC vivo with a nucleic acid, for treatment or prophylaxis of a condition
 CC caused in human or a non-human animal by a defect and/or a deficiency
 CC in a gene, immunisation and antisense therapy of a human or a non-human
 CC animal. It is useful for transfecting bronchial and lung epithelium and
 CC corneal endothelium for gene therapy for cystic fibrosis, asthma and
 CC also various cancers and viral infections for example human
 CC immunodeficiency virus (HIV) infection. It is also useful as a vaccine
 CC or for therapy of neuroblastoma and the effective transfection of
 CC primary smooth muscle cells, cardiac myocytes and hematopoietic cells.
 CC Haematopoietic cell transfection enables gene therapy, gene vaccination
 CC and antisense therapy of diseases involving hematopoietic cells,
 CC including leukaemia and bone marrow stem cell disorders, for example
 CC transfection of a cytokine gene may be used for adjuvant immunotherapy.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for example
 CC in glaucoma. A gene that prevents proliferation of cells in blood
 CC vessel walls is introduced using complex of the invention to reduce
 CC restenosis. The present sequence is cyclic integrin-binding peptide
 CC of the invention. This peptide is specific for alpha5beta1 integrin.
 CC
 XX Sequence 12 AA:
 SQ
 Query Match 80.6%; Score 50; DB 23; Length 12;
 Best Local Similarity 75.0%; Pred. No. 0.037;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 1 GAXXXETAWACG 12
 II |
 DB 1 GACRRETAWACG 12
 RESULT 5
 AAR76190
 ID AAR76190 standard; peptide: 13 AA.
 XX
 AC AAR76190;
 XX
 DT 24-JAN-1996 (first entry)
 XX
 DE Integrin binding peptide #2.
 XX
 KW High affinity; integrin binding peptide; alpha5/beta1; alpha/beta5;
 KW alpha/beta3; RGD; stable configuration; wound healing;
 KW osteoclast attachment; bone; angiogenesis; metastasis; tumour;
 KW smooth muscle cell migration.
 XX
 OS Synthetic.
 OS
 FH Key Location/Qualifiers
 FT Disulfide-bond 3.11
 FT
 PN WO9514714-A1.
 XX

PD 01-JUN-1995.
 XX
 PF 22-NOV-1994; 94WO-US13542.
 XX
 PR 04-AUG-1994; 94US-0286861.
 PR 24-NOV-1993; 93US-0158001.
 XX
 PA (LJOL-) LA JOLLA CANCER RES FOUND.
 PI
 PI Kolvunen E, Ruoslahti E;
 XX
 DR WPI; 1995-206899/27.
 XX
 PT High affinity integrin binding peptides - can be used to attach
 PT cells to a substrate, inhibit the attachment of osteoclasts to bone,
 PT promote wound healing, inhibit angiogenesis, metastasis of tumours
 PT and migration of smooth muscle cells
 XX
 PS Example 1; Page 24; 86pp; English.
 XX
 CC The sequences given in AAR76185-200 and AAR79073-94 are high affinity
 CC integrin binding peptides which bind to various integrins. Peptides
 CC which bind to alpha5/beta1 integrins contain the motifs given in
 CC AAR76185-86 and peptides which bind to alphaV/beta5 and alphaV/beta3
 CC integrins contain the motif given in AAR76187. AlphaV/beta5 integrins
 CC are also bound by RGD containing peptides. These peptides assume a
 CC conformationally stabilised configuration which is due to the
 CC formation of a disulphide bond, a peptide bond or a lactam bond.
 CC These peptides may be used for isolating the complementary integrin
 CC from a sample mixture by contacting them under ionic conditions to
 CC allow binding of the integrin to the peptide and then separating the
 CC integrin from the peptide. They can be used for attaching cells to
 CC a substrate, by binding them to the substrate with the cell. The
 CC peptides promote wound healing when applied locally and inhibit the
 CC attachment of osteoclasts to bone. They inhibit angiogenesis,
 CC metastasis of tumours and migration of smooth muscle cells.
 XX
 SO Sequence 13 AA:
 Query Match 80.6%; Score 50; DB 16; Length 13;
 Best Local Similarity 75.0%; Pred. No. 0.04;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 GAXXXETAMACG 12
 II |||||
 DB 1 GACRRRETAMACG 12
 RESULT 6
 AAB20796
 ID AAB20796 standard; peptide: 13 AA.
 XX
 AC AAB20796;
 XX
 DT 03-JAN-2001 (first entry)
 XX
 DE Alpha-5-beta-1 integrin binding peptide SEQ ID NO:63.
 XX
 XX Target protein binding element; protein level control; eukaryotic;
 KW ubiquitination recognition element; treatment; infection; cancer;
 KW inflammatory condition; genetic disease; insecticide; herbicide;
 KW antiviral; antiparasitic; hepatotropic; antineoplastic; cytostatic;
 KW tumour; pest control; pesticide; rodenticide; fungicide;
 KW gene expression; gene therapy.
 XX
 OS Unidentified.
 XX
 PN WO200047220-A1.
 XX
 PD 17-AUG-2000.
 XX
 PF 11-FEB-2000; 2000WO-US03436.
 XX

PR 12-FEB-1999; 99US-0119851.
 PR 28-SEP-1999; 99US-0406781.
 XX
 PA (PROT-) PROTELINX INC.
 XX
 PI Kenten JH, Roberts SF, Lebowitz MS;
 XX
 DR WPI; 2000-565258/52.
 XX
 PT Novel compounds for modulating the ubiquitination of target proteins
 PT comprising a ubiquitination recognition element-target protein element
 PT fusion, useful for treating viral infections -
 XX
 PS Disclosure; Page 22; 106pp; English.
 XX
 CC The present invention describes a compound (I) for activating the
 CC ubiquitination (Ub'n) of a target protein comprising a Ub'n recognition
 CC (peptide) element (URE) covalently linked to a target protein (peptide)
 CC element (TPE). (I) can have antiviral, antiparasitic, hepatotropic,
 CC antineoplastic and cytostatic activities. The compound of (I) may be
 CC used to treat a viral infection (especially hepatitis A, B, C or G, HIV-1
 CC or 2, Herpes, CMV, rabies or Rouse sarcoma virus (RSV)), parasitic
 CC infection, an infection caused by an eukaryotic organism in a mammal, to
 CC treat a tumour or to control pests. The compound may also be used to
 CC screen for target protein binding elements, to develop pesticides (e.g.
 CC insecticides, rodenticides, fungicides and herbicides) and to control
 CC gene expression (gene therapy). The present sequence represents an
 CC example of a target protein binding element given in the exemplification
 CC of the present invention.
 XX
 SO Sequence 13 AA:
 Query Match 80.6%; Score 50; DB 21; Length 13;
 Best Local Similarity 75.0%; Pred. No. 0.04;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 OY 1 GAXXXETAMACG 12
 II |||||
 DB 1 GACRRRETAMACG 12
 RESULT 7
 AAU74986
 ID AAU74986 standard; Peptide: 23 AA.
 XX
 AC AAU74986;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Transfection associated, integrin binding peptide #10.
 XX
 XX Cyclic; virucide; human immunodeficiency virus; HIV; cytostatic;
 KW ophthalmological; vasotropic; vaccine; gene therapy; transfection;
 KW cystic fibrosis; asthma; cancer; leukaemia; glaucoma; gene vaccination;
 KW anti-sense therapy; eye disease; corneal organ transplant; integrin;
 KW transfection; restenosis.
 XX
 OS Synthetic.
 XX
 PN Key Location/Qualifiers
 XX
 FT Disulphide_bond 1..9
 FT Disulphide_bond /note= "Cysteine residues linked by a disulfide
 FT bond to form a cyclic structure"
 FT Disulphide_bond 14..22
 FT Disulphide_bond /note= "Cysteine residues linked by a disulfide
 FT bond to form a cyclic structure"
 XX
 PN WO200192543-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-GB02396.
 XX

| | | | |
|----------|--|-----------------|---------------------------------|
| PR | 30-MAY-2000: | 2000GB-0013089. | |
| PR | 30-MAY-2000: | 2000GB-0013090. | |
| PR | 01-MAY-2001: | 2001US-287410P. | |
| PA | (ICHI-) | ICH PRODN LTD. | |
| PI | Hart SL: | | |
| DR | WPI: 2002-114355/15. | | |
| XX | | | |
| PT | Transfecting confluent cells with nucleic acid for gene therapy or gene | | |
| PT | vaccination, comprises contacting the cells with a receptor-targeted | | |
| PT | vector having the nucleic acid and an agent that disrupts cell-cell | | |
| PT | junctions - | | |
| XX | | | |
| PS | Claim 17; Page 17; 11pp; English. | | |
| XX | | | |
| CC | The invention describes transfecting (I) confluent cells or other slowly | | |
| CC | dividing or non-dividing cells that are in contact with each other, with | | |
| CC | a nucleic acid. The method comprises contacting the cells with a | | |
| CC | receptor-targeted vector comprising the nucleic acid, and an agent that | | |
| CC | disrupts cell-cell junctions under conditions suitable to effect | | |
| CC | transfection. (I) is useful for transfecting bronchial and lung | | |
| CC | epithelium for gene therapy for cystic fibrosis, asthma and also various | | |
| CC | cancers and viral infections e.g. human immunodeficiency virus (HIV) | | |
| CC | infection. Hematopoietic cell transfection enables gene therapy, gene | | |
| CC | vaccination and anti-sense therapy of diseases involving haematopoietic | | |
| CC | cells, including leukaemia and bone marrow stem cell disorders. | | |
| CC | Transfection of corneal endothelium is useful for treatment of eye | | |
| CC | disease affecting the cornea or corneal organ transplants, for e.g. in | | |
| CC | glaucoma. A gene preventing cell proliferation in blood vessel walls is | | |
| CC | introduced using an integrin targeting transfection vector complex (II) | | |
| CC | to reduce restenosis. (II) is useful for intracellular transport and | | |
| CC | delivery of anti-sense oligonucleotides, which enables antiviral and | | |
| CC | cancer therapy and is effective in transporting large DNA molecules. | | |
| CC | This sequence represents a peptide that will permit cyclisation by | | |
| CC | disulfide bond formation. It lacks the conserved RGD amino acid | | |
| CC | integrin binding sequence, but binds integrins, allowing the nucleic | | |
| CC | acid to pass into the cell, described in the method of the invention. | | |
| XX | | | |
| SQ | Sequence 23 AA: | | |
| QY | Query Match | 80.6%; | Score 50; DB: 23; Length 23; |
| | Best Local Similarity | 75.0%; | Precl. No. 0.07; |
| | Matches 9; Conservative | 0; | Mismatches 3; Indels 0; Gaps 0; |
| DB | 1 GAXXXETAMACG 12 | | |
| | | | |
| | 12 GACRRETAMACG 23 | | |
| RESULT 8 | | | |
| ID | AAE17116 | | |
| AC | AAE17116 standard; peptide; 23 AA. | | |
| XX | AAE17116: | | |
| DT | 18-APR-2002 (first entry) | | |
| XX | | | |
| DE | Cyclic integrin-binding peptide #18. | | |
| XX | | | |
| KM | Integrin binding component; polyanionic nucleic acid-binding component; | | |
| KM | 11pD component; prophylaxis; immunisation; anti-sense therapy; asthma; | | |
| KM | cystic fibrosis; cancer; viral infection; human immunodeficiency virus; | | |
| KM | HIV infection; vaccine; neuroblastoma; bone marrow stem cell disorder; | | |
| KM | leukaemia; adjuvant immunotherapy; eye disease; glaucoma; restenosis; | | |
| XX | integrin-binding peptide; cyclic. | | |
| OS | Unidentified. | | |
| XX | | | |
| XX | WO200192542-A2. | | |
| XX | | | |
| DD | 06-DEC-2001. | | |

| | | |
|----------|---|---|
| XX | PF | 30-MAY-2001; 2001WO-GS02394. |
| XX | PR | 30-MAY-2000; 2000GB-0013089. |
| XX | PR | 30-MAY-2000; 2000GB-0013090. |
| XX | PR | 01-MAY-2001; 2001US-287410P. |
| XX | PA | (ICHI-) ICH PRODN LTD. |
| XX | P1 | Hart SL; |
| XX | DR | WPI; 2002-139612/18. |
| PT | PT | Complex for transfecting cell with nucleic acid for treating, |
| PT | PT | preventing conditions caused by deficiency in a gene in humans, has |
| PT | PT | nucleic acid, lipid, integrin binding and polycationic nucleic |
| PS | PS | acid-binding components - |
| XX | XX | Claim 18; Page 78; 108pp; English. |
| CC | CC | The invention relates to integrin-targetting vectors having enhanced |
| CC | CC | transfection activity. The vector complex comprises a nucleic acid, |
| CC | CC | an integrin binding component, a polycationic nucleic acid-binding |
| CC | CC | component and a lipid component. The integrin binding component |
| CC | CC | comprises an integrin-binding element and a spacer element. Complex |
| CC | CC | of the invention is useful for transfecting cells in vitro or in |
| CC | CC | vivo with a nucleic acid, for treatment or prophylaxis of a condition |
| CC | CC | caused in human or a non-human animal by a defect and/or a deficiency |
| CC | CC | in a gene, immunisation and antisense therapy of a human or a non-human |
| CC | CC | animal. It is useful for transfecting bronchial and lung epithelium and |
| CC | CC | corneal endothelium for gene therapy for cystic fibrosis, asthma and |
| CC | CC | also various cancers and viral infections for example human |
| CC | CC | immunodeficiency virus (HIV) infection. It is also useful as a vaccine |
| CC | CC | or for therapy of neuroblastoma and the effective transfection of |
| CC | CC | primary smooth muscle cells, cardiac myocytes and haematopoietic cells. |
| CC | CC | Haematopoietic cell transfection enables gene therapy, gene vaccination |
| CC | CC | and antisense therapy of diseases involving haematopoietic cells, |
| CC | CC | including leukaemia and bone marrow stem cell disorders, for example |
| CC | CC | transfection of a cytokine gene may be used for adjuvant immunotherapy. |
| CC | CC | Transfection of corneal endothelium is useful for treatment of eye |
| CC | CC | disease affecting the cornea or corneal organ transplants, for example |
| CC | CC | in glaucoma. A gene that prevents proliferation of cells in blood |
| CC | CC | vessel walls is introduced using complex of the invention to reduce |
| CC | CC | restenosis. The present sequence is cyclic integrin-binding peptide |
| CC | CC | of the invention. |
| SO | SO | Sequence 23 AA: |
| OY | OY | 1 GAXXXETAWMACG 12 |
| DB | DB | 12 GCACRETFAMACG 23 |
| RESULT 9 | | |
| ID | AAW95416 | |
| XX | AAW95416 | standard; peptide: 25 AA. |
| XX | AAW95416; | |
| DT | 18-MAR-1999 | (first entry) |
| DE | Integrin-binding peptide. | |
| XX | Integrin; transfection complex; integrin-binding; lipid; immunisation; | |
| KM | antisense therapy; enzyme, therapeutic agent; immunogen; cystic fibrosis; | |
| KW | cancer; viral infection; human immunodeficiency virus; cardiovascular; | |
| XX | restenosis; leukaemia; asthma; glaucoma; cyclic; circular. | |
| XX | Synthetic. | |

| | | |
|-----------|---|------------------------------------|
| XX | Key | Location/Qualifiers |
| FT | Disulfide-bond | 3.11 |
| XX | | /note= "disulphide bridge" |
| PN | | |
| XX | MO9854347-A1. | |
| PD | | |
| XX | 03-DEC-1998. | |
| XX | | |
| XX | 29-MAY-1998; | 98WO-GB01577. |
| FR | 29-MAY-1997; | 97GB-0011115. |
| XX | | |
| PA | (CHIL-) INST CHILD HEALTH. | |
| XX | | |
| PI | Hart SL; | |
| XX | | |
| DR | WPI: 1999-045366/04. | |
| XX | | |
| PT | New integrin-targeting transfection complex including lipid - used | |
| XX | to improve transfection efficiency for a very wide range of cells, | |
| PT | useful in, e.g. antisense therapy and genetic immunisation | |
| XX | | |
| PS | Claim 12; Page 50; 70pp: English. | |
| XX | | |
| CC | The invention relates to an integrin-targeting transfection complex. The | |
| CC | complex comprises a nucleic acid, an integrin-binding component, a | |
| CC | polycationic nucleic acid-binding component and a lipid. The complexes | |
| CC | are used for in vivo or in vitro transfection of cells, specifically: | |
| CC | (i) for treatment or prevention of disease (in humans or other animals) | |
| CC | caused by defective or deficient genes; (ii) for immunisation; (iii) for | |
| CC | antisense therapy, and (iv) for protein production in host cells, e.g. | |
| CC | of enzymes, therapeutic agents, vaccinating immunogens and diagnostic | |
| CC | antigens. Typical of the diseases that can be treated or prevented are | |
| CC | cystic fibrosis, cancer, viral infection (e.g. human immunodeficiency | |
| CC | virus), cardiovascular disease (e.g. restenosis), leukaemia, asthma and | |
| CC | glaucoma. Incorporation of the lipid into the complex increases | |
| CC | transfection levels from 1-10 percent to over 50 percent. This effect is | |
| CC | observed with all cell types tested including those that are resistant to | |
| CC | transfection by most plasmid vectors. The complexes can carry large | |
| CC | genes, up to 125 kb, e.g. an artificial chromosome. The present sequence | |
| CC | represents a claimed example of an integrin-binding peptide used in the | |
| CC | transfection complexes. | |
| XX | | |
| SO | Sequence | 25 AA; |
| QY | Query Match | 80.6%; Score*50; DB 20; Length 25; |
| | Best Local Similarity | 75.0%; Pred. No. 0.075; |
| | Matches | 9; Conservative |
| | | 0; Mismatches |
| | | 3; Indels |
| | | 0; Gaps |
| | | 0; |
| DB | 1 GAXXETAMACG 12 | |
| | | |
| | 1 GACRRETAMACG 12 | |
| RESULT 10 | | |
| AAU74968 | | |
| ID | AAU74968 standard; Peptide: 25 AA. | |
| XX | AAU74968; | |
| XX | | |
| DT | 09-APR-2002 (first entry) | |
| XX | | |
| DE | Non-RGD containing alpha 5 beta 1 integrin binding peptide. | |
| KW | Cyclic; virucide; human immunodeficiency virus; HIV; cytostatic; | |
| KW | ophthalmological; vasotropic; vaccing; gene therapy; transfection; | |
| KW | cystic fibrosis; asthma; cancer; leukaemia; glaucoma; gene vaccination; | |
| KW | anti-sense therapy; eye disease; corneal organ transplant; integrin; | |
| XX | transfection; restenosis; alpha 5 beta 1 integrin. | |
| XX | | |
| CS | Synthetic. | |
| XX | | |

| Key | Location/Qualifiers |
|---|---|
| Disulfide-bond | 3..11 |
| Disulfide-bond | /note- "Cysteine residues linked by a disulphide bond to form a cyclic structure" |
| Disulfide-bond | 16..24 |
| Disulfide-bond | /note- "Cysteine residues linked by a disulphide bond to form a cyclic structure" |
| WO200192543-A2. | |
| 06-DEC-2001. | |
| 30-MAY-2001; 2001WO-GB02396. | |
| 30-MAY-2000; 2000GB-0013089. | |
| 30-MAY-2000; 2000GB-0013090. | |
| 01-MAY-2001; 2001US-287410P. | |
| (ICHI-) ICH PROD N LTD. | |
| Hart SL; | |
| WPI; 2002-114355/15. | |
| Transflecting confluent cells with nucleic acid for gene therapy or gene vaccination, comprises contacting the cells with a receptor-targeted vector having the nucleic acid and an agent that disrupts cell-cell junctions | |
| Disclosure; Page 15; 11pp; English. | |
| The invention describes transfecting (I) confluent cells or other slowly dividing or non-dividing cells that are in contact with each other, with a nucleic acid. The method comprises contacting the cells with a receptor-targeted vector comprising the nucleic acid, and an agent that disrupts cell-cell junctions under conditions suitable to effect transfection. (II) is useful for transfecting bronchial and lung epithelium for gene therapy for cystic fibrosis, asthma and also various cancers and viral infections e.g. human immunodeficiency virus (HIV) infection. Haematopoietic cell transfection enables gene therapy, gene vaccination and anti-sense therapy of diseases involving haematopoietic cells, including leukaemia and bone marrow stem cell disorders. | |
| Transfection of corneal endothelium is useful for treatment of eye disease affecting the cornea or corneal organ transplants, for e.g. in glaucoma. A gene preventing cell proliferation in blood vessel walls is introduced using an integrin targeting transfection vector complex (II) to reduce restenosis. (II) is useful for intracellular transport and delivery of anti-sense oligonucleotides, which enables antiviral and cancer therapy and is effective in transporting large DNA molecules. | |
| This sequence represents a cyclic peptide which lacks the conserved RGD integrin binding amino acid sequence but can bind integrins to allow the nucleic acid to pass into the cell, described in the method of the invention. | |
| Sequence | 25 AA: |
| Query Match | 80.6%: Score 50; DB 23; Length 25; |
| Best Local Similarity | 75.0%: Pred. NO. 0.075; 3; Indels 0; Gaps 0; |
| Matches | 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0; |
| 1 GAXXXETAWACG 12 | |
| 11 | |
| 1 GACRRETAWACG 12 | |
| RESULT 11 | |
| AAE17106 | |
| ID AAE17106 standard; peptide; 25 AA. | |
| AAE17106; | |
| 18-APR-2002 (first entry) | |

DE Integrin-binding peptide 9.
 XX
 KW Integrin binding component; polycationic nucleic acid-binding component;
 KW lipid component; prophylaxis; immunisation; antisense therapy; asthma;
 KW cystic fibrosis; cancer; viral infection; human immunodeficiency virus;
 KW HIV infection; vaccine; neuroblastoma; bone marrow stem cell disorder;
 KW leukaemia; adjuvant immunotherapy; eye disease; glaucoma; restenosis;
 KW Integrin-binding peptide.
 XX
 OS unidentified.
 XX
 PN WO200192542-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-GB02394.
 XX
 PR 30-MAY-2000; 2000GB-0013089.
 PR 30-MAY-2000; 2000GB-0013090.
 PR 01-MAY-2001; 2001US-287410P.
 XX
 PA (ICHI-) ICH PRODN LTD.
 XX
 PI Hart SL.
 DR WPI; 2002-139612/18.
 XX
 PT Complex for transfecting cell with nucleic acid for treating,
 PT preventing conditions caused by deficiency in a gene in humans, has
 PT nucleic acid, lipid, integrin binding and polycationic nucleic
 PT acid-binding components -
 XX
 PS Disclosure: Page 7; 108pp; English.
 XX
 CC The invention relates to Integrin-targeting vectors having enhanced
 CC transfection activity. The vector complex comprises a nucleic acid,
 CC an integrin binding component, a polycationic nucleic acid-binding
 CC component and a lipid component. The integrin binding component
 CC comprises an integrin-binding element and a spacer element. Complex
 CC of the invention is useful for transfecting cells in vitro or in
 CC vivo with a nucleic acid, for treatment or prophylaxis of a condition
 CC caused in human or a non-human animal by a defect and/or a deficiency
 CC in a gene, immunisation and antisense therapy of a human or a non-human
 CC animal. It is useful for transfecting bronchial and lung epithelium and
 CC corneal endothelium for gene therapy for cystic fibrosis, asthma and
 CC also various cancers and viral infections for example human
 CC immunodeficiency virus (HIV) infection. It is also useful as a vaccine
 CC or for therapy of neuroblastoma and the effective transfection of
 CC primary smooth muscle cells, cardiac myocytes and haematopoietic cells.
 CC Haematopoietic cell transfection enables gene therapy, gene vaccination
 CC and antisense therapy of diseases involving haematopoietic cells,
 CC including leukaemia and bone marrow stem cell disorders, for example
 CC transfection of a cytokine gene may be used for adjuvant immunotherapy.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for example
 CC in glaucoma. A gene that prevents proliferation of cells in blood
 CC vessel walls is introduced using complex of the invention to reduce
 CC restenosis. The present sequence is integrin-binding peptide
 CC of the invention. This peptide is specific for alphabeta1 integrin.
 XX
 SQ Sequence 25 AA;

Query Match 80.6%; Score 50; DB 23; Length 25;
 Best Local Similarity 75.0%; Pred. No. 0.075;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GAXXXETAWAG 12
 II |IIIIII|
 DB 1 GACRETAWACG 12

RESULT 12
 AAU74971

ID AAU74971 standard: Peptide: 28 AA.
 XX
 AC AAU74971;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Alpha 5 beta 1 Integrin binding oligo lysine peptide.
 XX
 KW Cyclic; virucide; human immunodeficiency virus; HIV; cytostatic;
 KW ophthalmological; vasotropic; vaccine; gene therapy; transfection;
 KW cystic fibrosis; asthma; cancer; leukaemia; glaucoma; gene vaccination;
 KW anti-sense therapy; eye disease; corneal organ transplant; integrin;
 KW transfection; restenosis; alpha 5 beta 1 Integrin.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Region 1..16
 FT /note= "Polycationic nucleic acid binding sequence"
 FT Peptide 17..28
 FT /note= "This sequence provides the alpha v beta 1
 FT Integrin binding specificity"
 FT Disulfide-bond 19..27
 FT /note= "Cysteine residues linked by a disulphide
 FT bond to form a cyclic structure"
 PN WO200192543-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-GB02396.
 XX
 PR 30-MAY-2000; 2000GB-0013089.
 PR 30-MAY-2000; 2000GB-0013090.
 PR 01-MAY-2001; 2001US-287410P.
 XX
 PA (ICHI-) ICH PRODN LTD.
 XX
 PI Hart SL.
 DR WPI; 2002-114355/15.
 XX
 PT Transfecting confluent cells with nucleic acid for gene therapy or gene
 PT vaccination, comprises contacting the cells with a receptor-targeted
 PT vector having the nucleic acid and an agent that disrupts cell-cell
 PT junctions -
 XX
 PS Example 5; Page 33; 111pp; English.
 XX
 CC The invention describes transfecting (I) confluent cells or other slowly
 CC dividing or non-dividing cells that are in contact with each other, with
 CC a nucleic acid. The method comprises contacting the cells with a
 CC receptor-targeted vector comprising the nucleic acid, and an agent that
 CC disrupts cell-cell junctions under conditions suitable to effect
 CC transfection. (II) is useful for transfecting bronchial and lung
 CC epithelium for gene therapy for cystic fibrosis, asthma and also various
 CC cancers and viral infections e.g. human immunodeficiency virus (HIV)
 CC infection. Haematopoietic cell transfection enables gene therapy, gene
 CC vaccination and anti-sense therapy of diseases involving haematopoietic
 CC cells, including leukaemia and bone marrow stem cell disorders.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for e.g. in
 CC glaucoma. A gene preventing cell proliferation in blood vessel walls is
 CC introduced using an integrin targeting transfection vector complex (II)
 CC to reduce restenosis. (III) is useful for intracellular transport and
 CC delivery of anti-sense oligonucleotides, which enables antiviral and
 CC cancer therapy and is effective in transporting large DNA molecules.
 CC This sequence represents a cyclic peptide that lacks the conserved RGD
 CC integrin binding amino acid sequence but can bind integrins to allow
 CC the nucleic acid to pass into the cell, described in the method of the
 CC invention.
 XX
 SQ Sequence 28 AA;

Query Match 80.6%; Score 50; DB 23; Length 28;
 Best Local Similarity 75.0%; Pred. No. 0.084;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GAXXXETAMACG 12
 || |||||
 DB 17 GACRRETAMACG 28

RESULT 13
 AAE17121
 ID AAE17121 standard; peptide: 28 AA.
 XX
 AC AAE17121;
 XX
 DT 18-APR-2002 (first entry)
 XX
 DE Integrin-binding oligolysine-peptide 6.
 XX
 KW Integrin binding component; polycationic nucleic acid-binding component;
 KW lipid component; prophylaxis; immunisation; antisense therapy; asthma;
 KW cystic fibrosis; cancer; viral infection; human immunodeficiency virus;
 KW HIV infection; vaccine; neuroblastoma; bone marrow stem cell disorder;
 KW leukaemia; adjuvant immunotherapy; eye disease; glaucoma; restenosis;
 KW Integrin-binding peptide.
 XX
 OS Unidentified.
 XX
 FH Key Location/Qualifiers
 FT Disulfide-bond 19..27
 FT /note= "Forms a cyclic structure"
 XX
 PN WO200192542-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-GB02394.
 XX
 PR 30-MAY-2000; 2000GB-0013089.
 PR 30-MAY-2000; 2000GB-0013090.
 PR 01-MAY-2001; 2001US-287410P.
 XX
 PA (ICHT-) ICH PRODN LTD.
 PI Hart SL;
 PI
 DR WPI: 2002-139612/18.
 XX
 PT Complex for transfecting cell with nucleic acid for treating,
 PT preventing conditions caused by deficiency in a gene in humans, has
 PT nucleic acid, lipid, integrin binding and polycationic nucleic
 PT acid-binding components -
 XX
 PS Example 13; Page 50; 108bp; English.

The invention relates to integrin-targeting vectors having enhanced transfection activity. The vector complex comprises a nucleic acid, an integrin binding component, a polycationic nucleic acid-binding component and a lipid component. The integrin binding component comprises an integrin-binding element and a spacer element. Complex of the invention is useful for transfecting cells in vitro or in vivo with a nucleic acid, for treatment or prophylaxis of a condition caused in human or a non-human animal by a defect and/or a deficiency in a gene, immunisation and antisense therapy of a human or a non-human animal. It is useful for transfecting bronchial and lung epithelium and corneal endothelium for gene therapy for cystic fibrosis, asthma and also various cancers and viral infections for example human immunodeficiency virus (HIV) infection. It is also useful as a vaccine or for therapy of neuroblastoma and the effective transfection of primary smooth muscle cells, cardiac myocytes and haematopoietic cells. Haematopoietic cell transfection enables gene therapy, gene vaccination and antisense therapy of diseases involving haematopoietic cells,

CC including leukaemia and bone marrow stem cell disorders, for example
 CC transfection of a cytokine gene may be used for adjuvant immunotherapy.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for example
 CC in glaucoma. A gene that prevents proliferation of cells in blood
 CC vessel walls is introduced using complex of the invention to reduce
 CC restenosis. The present sequence is integrin-binding oligolysine
 CC peptide of the invention.
 XX
 SQ Sequence 28 AA:

Query Match 80.6%; Score 50; DB 23; Length 28;
 Best Local Similarity 75.0%; Pred. No. 0.084;
 Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 1 GAXXXETAMACG 12
 || |||||
 DB 17 GACRRETAMACG 28

RESULT 14
 AAU74976
 ID AAU74976 standard; Peptide: 31 AA.
 XX
 AC AAU74976;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Integrin binding oligo lysine peptide #2.
 XX
 KW Cyclic; virucide; human immunodeficiency virus; HIV; cytostatic;
 KW ophthalmological; vasotropic; vaccine; gene therapy; transfection;
 KW cystic fibrosis; asthma; cancer; leukaemia; glaucoma; gene vaccination;
 KW anti-sense therapy; eye disease; corneal organ transplant; integrin;
 KW transfection; restenosis.
 XX
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Region 1..16
 FT /note= "Polycationic nucleic acid binding sequence"
 FT Region 17..21
 FT /label= Spacer
 FT Misc-difference 17
 FT /label= Unknown
 FT Misc-difference 19
 FT /label= Unknown
 FT Disulfide-bond 22..30
 FT /note= "Cysteine residues linked by a disulphide
 FT bond to form a cyclic structure"
 XX
 PN WO200192543-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 30-MAY-2001; 2001WO-GB02396.
 XX
 PR 30-MAY-2000; 2000GB-0013089.
 PR 30-MAY-2000; 2000GB-0013090.
 PR 01-MAY-2001; 2001US-287410P.
 XX
 PA (ICHT-) ICH PRODN LTD.
 PI Hart SL;
 PI
 DR WPI: 2002-114355/15.
 XX
 PT Transfecting confluent cells with nucleic acid for gene therapy or gene
 PT vaccination, comprises contacting the cells with a receptor-targeted
 PT vector having the nucleic acid and an agent that disrupts cell-cell
 PT junctions -
 XX
 PS Example 13; Page 55; 111bp; English.

XX The invention describes transfecting (I) confluent cells or other slowly
 CC dividing or non-dividing cells that are in contact with each other, with
 CC a nucleic acid. The method comprises contacting the cells with a
 CC receptor-targeted vector comprising the nucleic acid, and an agent that
 CC disrupts cell-cell junctions under conditions suitable to effect
 CC transfection. (I) is useful for transfecting bronchial and lung
 CC epithelium for gene therapy for cystic fibrosis, asthma and also various
 CC cancers and viral infections e.g. human immunodeficiency virus (HIV)
 CC infection. Haematopoietic cell transfection enables gene therapy, gene
 CC vaccination and anti-sense therapy of diseases involving haematopoietic
 CC cells, including leukaemia and bone marrow stem cell disorders.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for e.g. in
 CC glaucoma. A gene preventing cell proliferation in blood vessel walls is
 CC introduced using an integrin targeting transfection vector complex (II)
 CC to reduce restenosis. (II) is useful for intracellular transport and
 CC delivery of anti-sense oligonucleotides, which enables antiviral and
 CC cancer therapy and is effective in transporting large DNA molecules.
 CC This sequence represents a cyclic peptide that lacks the conserved RGD
 CC integrin binding amino acid sequence but binds to integrins to allow
 CC the nucleic acid to pass into the cell, described in the method of the
 CC invention.

SQ Sequence 31 AA;

Query Match 80.6%; Score 50; DB 23; Length 31;
 Best Local Similarity 75.0%; Pred. No. 0.093; Mismatches 3; Indels 0; Gaps 0;
 Matches 9; Conservative 0;

OY 1 GAXXXETAWACG 12
 II | | | | | | | |
 DB 20 GACRRETAWACG 31

RESULT 15
 AAU74982
 ID AAU74982 standard; Peptide; 10 AA.

AC AAU74982;

DT 09-APR-2002 (first entry)

DE Transfection associated, integrin binding peptide #6.

KW Cyclic; virucide; human immunodeficiency virus; HIV; cytostatic;

KW ophthalmological; vasotropic; vaccine; gene therapy; transfection;

KW cystic fibrosis; asthma; cancer; leukaemia; glaucoma; gene vaccination;

KW anti-sense therapy; eye disease; corneal organ transplant; integrin;

KW transfection; restenosis.

OS Synthetic.

FT Key location/Qualifiers

FT Disulphide_bond 1..9

FT /note= "Cysteine residues linked by a disulfide

FT bond to form a cyclic structure"

PN WO200192543-A2.

PD 06-DEC-2001.

PF 30-MAY-2001; 2001WO-CB02396.

PR 30-MAY-2000; 2000GB-0013089.

PR 30-MAY-2000; 2000GB-0013090.

PR 01-MAY-2001; 2001US-287410P.

XX (ICHI-) ICH PRODN LTD.

XX Hart SL.

XX WPI: 2002-114355/15.

XX Transfecting confluent cells with nucleic acid for gene therapy or gene
 PT vaccination, comprises contacting the cells with a receptor-targeted
 PT vector having the nucleic acid and an agent that disrupts cell-cell
 PT junctions
 PS Claim 17; Page 17; 11pp; English.

CC The invention describes transfecting (I) confluent cells or other slowly
 CC dividing or non-dividing cells that are in contact with each other, with
 CC a nucleic acid. The method comprises contacting the cells with a
 CC receptor-targeted vector comprising the nucleic acid, and an agent that
 CC disrupts cell-cell junctions under conditions suitable to effect
 CC transfection. (I) is useful for transfecting bronchial and lung
 CC epithelium for gene therapy for cystic fibrosis, asthma and also various
 CC cancers and viral infections e.g. human immunodeficiency virus (HIV)
 CC infection. Haematopoietic cell transfection enables gene therapy, gene
 CC vaccination and anti-sense therapy of diseases involving haematopoietic
 CC cells, including leukaemia and bone marrow stem cell disorders.
 CC Transfection of corneal endothelium is useful for treatment of eye
 CC disease affecting the cornea or corneal organ transplants, for e.g. in
 CC glaucoma. A gene preventing cell proliferation in blood vessel walls is
 CC introduced using an integrin targeting transfection vector complex (II)
 CC to reduce restenosis. (II) is useful for intracellular transport and
 CC delivery of anti-sense oligonucleotides, which enables antiviral and
 CC cancer therapy and is effective in transporting large DNA molecules.
 CC This sequence represents a peptide that will permit cyclisation by
 CC disulfide bond formation. It lacks the conserved RGD amino acid
 CC integrin binding sequence but binds integrins, allowing the nucleic
 CC acid to pass into the cell, described in the method of the invention.

SQ Sequence 10 AA;

Query Match 71.0%; Score 44; DB 23; Length 10;
 Best Local Similarity 100.0%; Pred. No. 0.33; Mismatches 0; Indels 0; Gaps 0;
 Matches 7; Conservative 0;

OY 6 ETAWACG 12
 | | | | | | | |
 DB 4 ETAWACG 10

Search completed: March 25, 2003, 08:19:41
 Job time: 41.7273 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:21:42 ; Search time 14.8485 Seconds
(Without alignments)
27.742 Million cell updates/sec

Title: US-09-646-532b-2
Perfect score: 62
Sequence: 1 GAXXXETAMACGA 14

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTU05.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------------------|-------------------|
| 1 | 50 | 80.6 | 12 | 4 | US-09-424-656-12 | Sequence 12, Appl |
| 2 | 50 | 80.6 | 13 | 1 | US-08-425-238-3 | Sequence 3, Appl1 |
| 3 | 50 | 80.6 | 13 | 2 | US-08-286-861-6 | Sequence 6, Appl1 |
| 4 | 50 | 80.6 | 13 | 4 | US-09-406-781-63 | Sequence 63, Appl |
| 5 | 50 | 80.6 | 25 | 4 | US-09-424-656-16 | Sequence 16, Appl |
| 6 | 44 | 71.0 | 12 | 4 | US-08-424-656-11 | Sequence 11, Appl |
| 7 | 40 | 64.5 | 13 | 1 | US-08-425-238-11 | Sequence 11, Appl |
| 8 | 38 | 61.3 | 9 | 1 | US-08-425-238-6 | Sequence 6, Appl1 |
| 9 | 38 | 61.3 | 9 | 2 | US-08-717-169-18 | Sequence 18, Appl |
| 10 | 38 | 61.3 | 9 | 2 | US-08-286-861-12 | Sequence 12, Appl |
| 11 | 38 | 61.3 | 9 | 4 | US-09-424-656-13 | Sequence 13, Appl |
| 12 | 38 | 61.3 | 13 | 3 | US-08-701-124-35 | Sequence 35, Appl |
| 13 | 38 | 61.3 | 13 | 3 | US-09-130-225-35 | Sequence 35, Appl |
| 14 | 38 | 61.3 | 13 | 4 | US-09-425-061-35 | Sequence 35, Appl |
| 15 | 35 | 56.5 | 237 | 4 | US-09-111-470-2 | Sequence 2, Appl1 |
| 16 | 35 | 56.5 | 662 | 4 | US-08-779-814-5 | Sequence 5, Appl1 |
| 17 | 35 | 56.5 | 928 | 3 | US-09-320-878-13 | Sequence 13, Appl |
| 18 | 35 | 56.5 | 928 | 4 | US-09-105-537-41 | Sequence 41, Appl |
| 19 | 35 | 56.5 | 1130 | 2 | US-08-519-547A-6 | Sequence 6, Appl1 |
| 20 | 35 | 56.5 | 3165 | 2 | US-08-459-146-3 | Sequence 3, Appl1 |
| 21 | 35 | 56.5 | 3165 | 2 | US-08-459-065-3 | Sequence 3, Appl1 |
| 22 | 34 | 54.8 | 222 | 3 | US-08-651-136C-14 | Sequence 14, Appl |
| 23 | 34 | 54.8 | 222 | 4 | US-09-229-911A-14 | Sequence 14, Appl |
| 24 | 34 | 54.8 | 294 | 3 | US-08-651-136C-24 | Sequence 24, Appl |
| 25 | 34 | 54.8 | 294 | 4 | US-09-229-911A-24 | Sequence 24, Appl |
| 26 | 33 | 53.2 | 170 | 4 | US-09-199-657A-299 | Sequence 299, App |
| 27 | 33 | 53.2 | 376 | 4 | US-09-135-020-113 | Sequence 113, App |

| | | | | | | |
|----|----|------|-----|---|--------------------|-------------------|
| 28 | 33 | 53.2 | 376 | 4 | US-09-135-010A-113 | Sequence 113, App |
| 29 | 33 | 53.2 | 376 | 4 | US-09-444-871-113 | Sequence 113, App |
| 30 | 33 | 53.2 | 376 | 4 | US-09-597-735-113 | Sequence 113, App |
| 31 | 33 | 53.2 | 376 | 4 | US-09-444-295-113 | Sequence 113, App |
| 32 | 33 | 53.2 | 376 | 4 | US-09-597-732-113 | Sequence 113, App |
| 33 | 33 | 53.2 | 430 | 4 | US-09-105-058C-21 | Sequence 21, Appl |
| 34 | 33 | 53.2 | 462 | 4 | US-09-129-112-2 | Sequence 2, Appl1 |
| 35 | 33 | 53.2 | 570 | 4 | US-09-135-020-114 | Sequence 114, App |
| 36 | 33 | 53.2 | 570 | 4 | US-09-135-010A-114 | Sequence 114, App |
| 37 | 33 | 53.2 | 570 | 4 | US-09-444-871-114 | Sequence 114, App |
| 38 | 33 | 53.2 | 570 | 4 | US-09-597-735-114 | Sequence 114, App |
| 39 | 33 | 53.2 | 570 | 4 | US-09-444-295-114 | Sequence 114, App |
| 40 | 33 | 53.2 | 570 | 4 | US-09-597-732-114 | Sequence 114, App |
| 41 | 33 | 53.2 | 581 | 4 | US-09-135-021-80 | Sequence 80, Appl |
| 42 | 33 | 53.2 | 581 | 4 | US-09-135-010A-116 | Sequence 116, App |
| 43 | 33 | 53.2 | 581 | 4 | US-09-597-735-116 | Sequence 116, App |
| 44 | 33 | 53.2 | 581 | 4 | US-09-597-732-116 | Sequence 116, App |
| 45 | 33 | 53.2 | 605 | 4 | US-09-105-058C-24 | Sequence 24, Appl |

ALIGNMENTS

```
RESULT 1
US-09-424-656-12
: Sequence 12, Application US/09424656
: Patent No. 6458026
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
: TITLE OF INVENTION: ENHANCED TRANSFECTION ACTIVITY
: NUMBER OF SEQUENCES: 16
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/424,656
: FILING DATE:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: GB 9711115.7
: FILING DATE: 29-MAY-1997
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 12 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-09-424-656-12

Query Match      80.6%; Score 50; DB 4; Length 12;
Best Local Similarity 75.0%; Pred. No. 0.011;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
DB 1 GACREITAMACG 12

RESULT 2
US-08-425-238-3
: Sequence 3, Application US/08425238
: Patent No. 5627263
: GENERAL INFORMATION:
: APPLICANT: Ruoslahti, Erkki
: APPLICANT: Koivunen, Erkki
: TITLE OF INVENTION: No. 5627263el Integrin-Binding Peptides
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
```

CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,238
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,001
FILING DATE: 24-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9775
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-425-238-3

Query Match 80.6%; Score 50; DB 1; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAXXXETAMACG 12
|| |||||
Db 1 GACRRETAMACG 12

RESULT 3
US-08-286-861-6
Sequence 6, Application US/08286861
Patent No. 5981478
GENERAL INFORMATION:
APPLICANT: Ruoslahti, Erkki
APPLICANT: Koivunen, Erkki
TITLE OF INVENTION: No. 5981478el Integrin-Binding Peptides
NUMBER OF SEQUENCES: 46
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/286,861
FILING DATE: 04-AUG-1994
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,001
FILING DATE: 24-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9992
TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-08-286-861-6

Query Match 80.6%; Score 50; DB 2; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||
Db 1 GACRRETAMACG 12

RESULT 4
US-09-406-781-63
Sequence 63, Application US/09406781
Patent No. 6306663
GENERAL INFORMATION:
APPLICANT: Keenen, John
APPLICANT: Roberts, Steven
TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUKARYOTIC ORGANISMS
FILE REFERENCE: 2757-3
CURRENT APPLICATION NUMBER: US/09/406,781
CURRENT FILING DATE: 1999-09-28
EARLIER APPLICATION NUMBER: 60/119,851
EARLIER FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 67
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 63
LENGTH: 13
TYPE: PRT
ORGANISM: Unknown organism
FEATURE:
OTHER INFORMATION: Description of Unknown Organism: binding peptide
US-09-406-781-63

Query Match 80.6%; Score 50; DB 4; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||
Db 1 GACRRETAMACG 12

RESULT 5
US-09-424-656-16
Sequence 16, Application US/09424656
Patent No. 6458026
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
TITLE OF INVENTION: ENHANCED TRANSECTON ACTIVITY
NUMBER OF SEQUENCES: 16
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30 (ERO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/424,656
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9711115.7
FILING DATE: 29-MAY-1997
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: circular
MOLECULE TYPE: peptide
US-09-424-656-16

Query Match 80.6%; Score 50; DB 4; Length 25;
Best Local Similarity 75.0%; Pred. No. 0.022;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
||| |||||
DB 1 GACRRETAMACG 12

RESULT 6
US-09-424-656-11
Sequence 11, Application US/09424656
Patent No. 6458026
GENERAL INFORMATION:

APPLICANT:
TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
TITLE OF INVENTION: ENHANCED TRANSFECTION ACTIVITY
NUMBER OF SEQUENCES: 16
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/424,656
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 971115.7
FILING DATE: 29-MAY-1997
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: circular
MOLECULE TYPE: peptide
US-09-424-656-11

Query Match 71.0%; Score 44; DB 4; Length 12;
Best Local Similarity 72.7%; Pred. No. 0.11;
Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMAC 11
||| |||||
DB 1 GACRRETAMAC 11

RESULT 7
US-08-425-238-11
Sequence 11, Application US/08425238
Patent No. 5627263
GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki
APPLICANT: Koivinen, Erkki
TITLE OF INVENTION: No. 5627263el Integrin-Binding Peptides
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,238
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,001
FILING DATE: 24-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 9775
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

FEATURE:
NAME/KEY: Peptide
LOCATION: 3
OTHER INFORMATION: /note="Xaa = an amino acid capable
of forming a disulfide bond."
FEATURE:
NAME/KEY: Peptide
LOCATION: 11
OTHER INFORMATION: /note="Xaa = an amino acid capable
of forming a disulfide bond." ;

US-08-425-238-11
Query Match 64.5%; Score 40; DB 1; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.56;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
||| |||||
DB 1 GACRRETAMACG 12

RESULT 8
US-08-425-238-6
Sequence 6, Application US/08425238
Patent No. 5627263
GENERAL INFORMATION:

APPLICANT: Ruoslahti, Erkki
APPLICANT: Koivinen, Erkki
TITLE OF INVENTION: No. 5627263el Integrin-Binding Peptides
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/425,238
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,001
FILING DATE: 24-NOV-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815

```

: REFERENCE/DOCKET NUMBER: P-LA 9775
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-425-238-6

Query Match      61.3%; Score 38; DB 1; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
DB 4 ETAMAC 9

RESULT 9
US-08-717-169-18
: Sequence 18, Application US/08717169
: Patent No. 5922676
: GENERAL INFORMATION:
: APPLICANT: Pasqualini, Renata
: APPLICANT: Ruoslahti, Erkki
: TITLE OF INVENTION: Methods of Inhibiting Angiogenesis and
: TITLE OF INVENTION: Ameliorating Cancer By Using Superfibronection
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell & Flores LLP
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/717,169
: FILING DATE: 20-SEP-1996
: CLASSIFICATION: 514
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn A.
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LJ 2017
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 18:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: peptide
: US-08-717-169-18

Query Match      61.3%; Score 38; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
DB 4 ETAMAC 9

RESULT 10
```

```

: US-08-286-861-12
: Sequence 12, Application US/08286861
: Patent No. 5981478
: GENERAL INFORMATION:
: APPLICANT: Ruoslahti, Erkki
: APPLICANT: Koivunen, Erkki
: TITLE OF INVENTION: No. 5981478e1 Integrin-Binding Peptides
: NUMBER OF SEQUENCES: 46
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Campbell and Flores
: STREET: 4370 La Jolla Village Drive, Suite 700
: CITY: San Diego
: STATE: California
: COUNTRY: USA
: ZIP: 92122
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/286,861
: FILING DATE: 04-AUG-1994
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/158,001
: FILING DATE: 24-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Campbell, Cathryn
: REGISTRATION NUMBER: 31,815
: REFERENCE/DOCKET NUMBER: P-LA 9992
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 535-9001
: TELEFAX: (619) 535-8949
: INFORMATION FOR SEQ ID NO: 12:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9 amino acids
: TYPE: amino acid
: TOPOLOGY: circular
: US-08-286-861-12

Query Match      61.3%; Score 38; DB 2; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
DB 4 ETAMAC 9

RESULT 11
US-09-424-656-13
: Sequence 13, Application US/09424656
: Patent No. 6458026
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
: TITLE OF INVENTION: ENHANCED TRANSFECTION ACTIVITY
: NUMBER OF SEQUENCES: 16
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO).
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/424,656
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: GB 9711115.7
: FILING DATE: 29-MAY-1997
: INFORMATION FOR SEQ ID NO: 13:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 9 amino acids
```

TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: circular
MOLECULE TYPE: peptide
US-09-424-656-13

Query Match 61.3%; Score 38; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
111111
DB 4 ETAMAC 9

RESULT 12
US-08-701-124-35
Sequence 35, Application US/08701124
Patent No. 5846782
GENERAL INFORMATION:
APPLICANT: Wickham, Thomas J.
APPLICANT: Roelivink, Petrus W.
TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF
TITLE OF INVENTION: CONSTRAINED PEPTIDE MOTIFS
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza - 49th Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/701,124
FILING DATE: 21-AUG-1996
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-701-124-35

Query Match 61.3%; Score 38; DB 2; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
111111
DB 6 ETAMAC 11

RESULT 13
US-09-130-225-35
Sequence 35, Application US/09130225
Patent No. 6057155
GENERAL INFORMATION:
APPLICANT: Wickham, Thomas J.
APPLICANT: Roelivink, Petrus W.
APPLICANT: Kovesdi, Imre
TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF
TITLE OF INVENTION: CONSTRAINED PEPTIDE MOTIFS
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza - 49th Floor
CITY: Chicago

STATE: Illinois
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/130,225
FILING DATE:
PRIOR APPLICATION NUMBER: US 8-701124
APPLICATION NUMBER: US 8-701124
FILING DATE: 21-AUG-1996
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-130-225-35

Query Match 61.3%; Score 38; DB 3; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11
111111
DB 6 ETAMAC 11

RESULT 14
US-09-455-061-35
Sequence 35, Application US/09455061
Patent No. 6329190
GENERAL INFORMATION:
APPLICANT: Wickham, Thomas J.
APPLICANT: Roelivink, Petrus W.
APPLICANT: Kovesdi, Imre
TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF
TITLE OF INVENTION: CONSTRAINED PEPTIDE MOTIFS
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Volt & Mayer, Ltd.
STREET: Two Prudential Plaza - 49th Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/455,061
FILING DATE: 06-DEC-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 9-130225
FILING DATE: 06-AUG-1998
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER: US 8-701124
FILING DATE: 21-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hefner, M. Daniel
REGISTRATION NUMBER: 41,826
REFERENCE/DOCKET NUMBER: 203128
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide

US-09-455-061-35

Query Match 61.3%; Score 38; DB 4; Length 13;
Best Local Similarity 100.0%; Pred. No. 1,2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11

Db 6 ETAMAC 11

RESULT 15

US-09-111-470-2

; Sequence 2, Application US/09111470
; Patent No. 6277959

; GENERAL INFORMATION:

; APPLICANT: Valladeau, Jenny

; APPLICANT: Ravel, Odile

; APPLICANT: Bates, Elizabeth E.M.

; APPLICANT: Ford, John

; APPLICANT: Saeland, Sem

; APPLICANT: Lebecque, Serge J.E.

; TITLE OF INVENTION: Mammalian Membrane Protein Genes;

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: DNAX Research Institute

; STREET: 901 California Avenue

; CITY: Palo Alto

; STATE: California

; COUNTRY: USA

; ZIP: 94304-1104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/111,470

; FILING DATE: 08-JUL-1998

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 60/053,080

; FILING DATE: 09-JUL-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ching, Edwin P.

; REGISTRATION NUMBER: 34,090

; REFERENCE/DOCKET NUMBER: SF0695

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (650)852-9196

; TELEFAX: (650)496-1200

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 237 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-09-111-470-2

Query Match 56.5%; Score 35; DB 4; Length 237;
Best Local Similarity 83.3%; Pred. No. 61;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAMAC 11

Db 100 ETAMSC 105

Search completed: March 25, 2003, 08:22:59
Job time : 15.8485 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 25, 2003, 08:22:27 ; Search time 13.1515 seconds
(without alignments)
56.911 Million cell updates/sec

Title: US-09-646-532B-2

Perfect score: 62

Sequence: 1 GAXXXETRWACGXA 14

Scoring table:

BLASTSUM62
Gapop 10.0 , Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCIT_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCITUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description |
|------------|-------|---------|--------------|-------|---------------------|
| 1 | 50 | 80.6 | 12 | 9 | US-10-137-435-12 |
| 2 | 50 | 80.6 | 13 | 9 | US-09-880-132-63 |
| 3 | 50 | 80.6 | 13 | 10 | US-09-364-597A-6 |
| 4 | 50 | 80.6 | 13 | 10 | US-09-880-149-63 |
| 5 | 50 | 80.6 | 23 | 9 | US-10-137-435-16 |
| 6 | 44 | 71.0 | 12 | 9 | US-10-137-435-11 |
| 7 | 38 | 61.3 | 9 | 9 | US-10-137-435-13 |
| 8 | 38 | 61.3 | 9 | 10 | US-09-364-597A-12 |
| 9 | 38 | 61.3 | 13 | 10 | US-09-969-192-35 |
| 10 | 37 | 59.7 | 38 | 10 | US-09-864-761-38378 |
| 11 | 36 | 58.1 | 138 | 9 | US-09-883-152-4 |
| 12 | 36 | 58.1 | 166 | 9 | US-09-747-835A-56 |
| 13 | 36 | 58.1 | 310 | 10 | US-09-886-055-485 |
| 14 | 36 | 58.1 | 2213 | 9 | US-10-184-644-549 |
| 15 | 35 | 56.5 | 202 | 10 | US-09-280-197-11 |
| 16 | 35 | 56.5 | 237 | 9 | US-09-862-802-2 |
| 17 | 35 | 56.5 | 237 | 9 | US-09-870-759-49 |
| 18 | 35 | 56.5 | 243 | 10 | US-09-764-870-464 |
| 19 | 35 | 56.5 | 246 | 10 | US-09-764-870-303 |

| | | | | | | |
|----|----|------|------|----|---------------------|-------------------|
| 20 | 35 | 56.5 | 830 | 10 | US-09-064-199-19 | Sequence 19, Appl |
| 21 | 35 | 56.5 | 928 | 9 | US-09-860-846-41 | Sequence 41, Appl |
| 22 | 35 | 56.5 | 928 | 10 | US-09-861-289-41 | Sequence 41, Appl |
| 23 | 35 | 56.5 | 1092 | 9 | US-09-423-126-5 | Sequence 5, Appl1 |
| 24 | 35 | 56.5 | 1106 | 10 | US-09-064-199-17 | Sequence 17, Appl |
| 25 | 35 | 56.5 | 1130 | 9 | US-10-104-595-6 | Sequence 6, Appl1 |
| 26 | 35 | 56.5 | 1130 | 10 | US-09-064-199-18 | Sequence 18, Appl |
| 27 | 35 | 56.5 | 1207 | 10 | US-09-064-199-16 | Sequence 16, Appl |
| 28 | 34 | 54.8 | 90 | 10 | US-09-864-761-37160 | Sequence 17160, A |
| 29 | 34 | 54.8 | 221 | 9 | US-09-738-626-4178 | Sequence 4178, Ap |
| 30 | 33 | 53.2 | 41 | 10 | US-09-158-180-5 | Sequence 5, Appl1 |
| 31 | 33 | 53.2 | 87 | 9 | US-09-764-868-1134 | Sequence 1134, Ap |
| 32 | 33 | 53.2 | 89 | 10 | US-09-864-761-35769 | Sequence 35769, A |
| 33 | 33 | 53.2 | 170 | 9 | US-09-975-179-299 | Sequence 299, App |
| 34 | 33 | 53.2 | 237 | 10 | US-09-746-284-1 | Sequence 1, Appl1 |
| 35 | 33 | 53.2 | 404 | 10 | US-09-815-242-11242 | Sequence 11242, A |
| 36 | 33 | 53.2 | 430 | 9 | US-10-128-870-21 | Sequence 21, Appl |
| 37 | 33 | 53.2 | 430 | 9 | US-10-131-685-21 | Sequence 21, Appl |
| 38 | 33 | 53.2 | 462 | 10 | US-09-129-112-2 | Sequence 2, Appl1 |
| 39 | 33 | 53.2 | 481 | 10 | US-09-158-180-2 | Sequence 2, Appl1 |
| 40 | 33 | 53.2 | 605 | 9 | US-10-128-870-24 | Sequence 24, Appl |
| 41 | 33 | 53.2 | 605 | 9 | US-10-131-685-24 | Sequence 24, Appl |
| 42 | 33 | 53.2 | 676 | 10 | US-09-840-125-2 | Sequence 2, Appl1 |
| 43 | 33 | 53.2 | 676 | 10 | US-09-813-148-3 | Sequence 3, Appl1 |
| 44 | 33 | 53.2 | 725 | 9 | US-09-764-868-755 | Sequence 755, App |
| 45 | 33 | 53.2 | 1114 | 9 | US-10-184-644-271 | Sequence 271, App |

ALIGNMENTS

RESULT 1
US-10-137-435-12
Sequence 12, Application US/10137435
Publication No. US20030013644A1
GENERAL INFORMATION:
APPLICANT: Institute of Child Health
TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
ENHANCED TRANSFECTION ACTIVITY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
STREET: c/o Institute of Child Health,
30 Gulliford street
CITY: London
COUNTRY: G.B.
ZIP: WC1N 1EH
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/137,435
FILING DATE: 03-May-2002
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/424,656
FILING DATE: <Unknown>
APPLICATION NUMBER: GB 9711115.7
FILING DATE: 29-May-1997
ATTORNEY/AGENT INFORMATION:
NAME: Hart, Stephen Lewis
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 12:
LENGTH: 12 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-137-435-12
Query Match 80.6%; Score 50; DB 9; Length 12;
Best Local Similarity 75.0%; Pred. No. 0.014;

Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||

Db 1 GACRRETAMACG 12

RESULT 2
US-09-880-132-63
; Sequence 63, Application US/09880132
; Patent No. US20020173049A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John
; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUKARYOTIC ORGANISMS
; FILE REFERENCE: 2757-6
; CURRENT APPLICATION NUMBER: US/09/880,132
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 09/406,781
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: 60/119,851
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 63
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: binding peptide
US-09-880-132-63

Query Match 80.6%; Score 50; DB 9; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.015;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||

Db 1 GACRRETAMACG 12

RESULT 3
US-09-364-597A-6
; Sequence 6, Application US/09364597A
; Patent No. US20020103130A1
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Kolvinen, Erkki
; TITLE OF INVENTION: No. US20020103130A1 Integrin-Binding Peptides
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/364,597A
; FILING DATE: 30-JUL-1999
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/158,001
; FILING DATE: 24-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/286,861
; FILING DATE: 04-AUG-1994
; ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 3419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (858) 535-9001
TELEFAX: (858) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-364-597A-6

Query Match 80.6%; Score 50; DB 10; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.015;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||

Db 1 GACRRETAMACG 12

RESULT 4
US-09-880-149-63
; Sequence 63, Application US/09880149
; Patent No. US20020146843A1
; GENERAL INFORMATION:
; APPLICANT: Kenten, John
; TITLE OF INVENTION: CONTROLLING PROTEIN LEVELS IN EUKARYOTIC ORGANISMS
; FILE REFERENCE: 2757-5
; CURRENT APPLICATION NUMBER: US/09/880,149
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: 09/406,781
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: 60/119,851
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 63
; LENGTH: 13
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURE:
; OTHER INFORMATION: Description of Unknown Organism: binding peptide
US-09-880-149-63

Query Match 80.6%; Score 50; DB 10; Length 13;
Best Local Similarity 75.0%; Pred. No. 0.015;
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
|| |||||

Db 1 GACRRETAMACG 12

RESULT 5
US-10-137-435-16
; Sequence 16, Application US/10137435
; Publication No. US20030013644A1
; GENERAL INFORMATION:
; APPLICANT: Institute of Child Health
; TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
; ENHANCED TRANSESECTION ACTIVITY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; STREET: C/O Institute of Child Health,
; 30 Guildford Street
; CITY: London
; COUNTRY: G.B.
; ZIP: WC1N 1EH
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/137,435
; FILING DATE: 03-May-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/424,656
; FILING DATE: <unknown>
; APPLICATION NUMBER: GB 9711115.7
; FILING DATE: 29-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Stephen Lewis
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-10-137-435-16
;
; Query Match
; Best Local Similarity 80.6%; Score 50; DB 9; Length 25;
; Match 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAXXETAWACG 12
DB 1 GACRRETAWACG 12
;
; RESULT 6
; US-10-137-435-11
; Sequence 11, Application US/10137435
; Publication No. US20030013644A1
; GENERAL INFORMATION:
; APPLICANT: Institute of Child Health
; TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
; ENHANCED TRANSECTIONS ACTIVITY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; STREET: c/o Institute of Child Health,
; 30 Gildford Street
; CITY: London
; COUNTRY: G.B.
; ZIP: WC1N 1EH
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/137,435
; FILING DATE: 03-May-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/424,656
; FILING DATE: <unknown>
; APPLICATION NUMBER: GB 9711115.7
; FILING DATE: 29-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Stephen Lewis
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-137-435-11
;
; Query Match
; 71.0%; Score 44; DB 9; Length 12;
```

```

;
; Best Local Similarity 72.7%; Pred. No. 0.15;
; Matches 8; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 GAXXETAWAC 11
DB 1 GACRRETAWAC 11
;
; RESULT 7
; US-10-137-435-13
; Sequence 13, Application US/10137435
; Publication No. US20030013644A1
; GENERAL INFORMATION:
; APPLICANT: Institute of Child Health
; TITLE OF INVENTION: INTEGRIN-TARGETING VECTORS HAVING
; ENHANCED TRANSECTIONS ACTIVITY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; STREET: c/o Institute of Child Health,
; 30 Gildford Street
; CITY: London
; COUNTRY: G.B.
; ZIP: WC1N 1EH
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/137,435
; FILING DATE: 03-May-2002
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/424,656
; FILING DATE: <unknown>
; APPLICATION NUMBER: GB 9711115.7
; FILING DATE: 29-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hart, Stephen Lewis
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: circular
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 13:
US-10-137-435-13
;
; Query Match
; 61.3%; Score 38; DB 9; Length 9;
; Best Local Similarity 100.0%; Pred. No. 26+05;
; Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 ETAWAC 11
DB 4 ETAWAC 9
;
; RESULT 8
; US-09-364-597A-12
; Sequence 12, Application US/09364597A
; Patent No. US20020103130A1
; GENERAL INFORMATION:
; APPLICANT: Ruoslahti, Erkki
; APPLICANT: Koivunen, Erkki
; TITLE OF INVENTION: No. US20020103130A1 Integrin-Binding Peptides
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
```

COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/364,597A
FILING DATE: 30-JUL-1999
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/158,001
FILING DATE: 24-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/286,861
FILING DATE: 04-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-LA 3419
TELECOMMUNICATION INFORMATION:
TELEPHONE: (858) 535-9001
TELEFAX: (858) 535-8949
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
TOPOLOGY: circular
US-09-364-597A-12

Query Match 61.3%; Score 38; DB 10; Length 9;
Best Local Similarity 100.0%; Pred. No. 2e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAWAC 11
|||||
DB 4 ETAWAC 9

RESULT 9
US-09-969-192-35
Sequence 35, Application US/09969192
Patent No. US20020151027A1
GENERAL INFORMATION:
APPLICANT: WICKHAM, THOMAS J.
ROELVINK, PETRUS W.
KOVESDI, IMRE
TITLE OF INVENTION: TARGETING ADENOVIRUS WITH USE OF
CONSTRAINED PEPTIDE MOTIFS
NUMBER OF SEQUENCES: 80
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza - 49th Floor
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/969,192
FILING DATE: 01-Oct-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 9-455061
FILING DATE: 06-DEC-1999
APPLICATION NUMBER: US 9-130225
FILING DATE: 06-AUG-1998
APPLICATION NUMBER: US 8-701124
FILING DATE: 21-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hefner, M. Daniel

REGISTRATION NUMBER: 41,826
REFERENCE/DOCKET NUMBER: 213564
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 13 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 35:
US-09-969-192-35

Query Match 61.3%; Score 38; DB 10; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 ETAWAC 11
|||||
DB 6 ETAWAC 11

RESULT 10
US-09-864-761-38378
Sequence 38378, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecm1ca-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 38378

```

; LENGTH: 38
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC005630.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 7.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 8.7
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.3
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 7.5
; OTHER INFORMATION: EST_HUMAN HIT: AW162304.1, EVALU6 3.00e-05
US-09-864-761-38378

Query Match          59.7%; Score 37; DB 10; Length 38;
Best Local Similarity 50.0%; Pred. No. 5.9;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 1 GAXXXETAMACGX 14
Db 24 GAVPQGAWMCQA 37

RESULT 11
US-09-883-152-4
; Sequence 4, Application US/09883152
; Publication No. US20030008284A1
; GENERAL INFORMATION:
; APPLICANT: Kennedy, Giulia
; APPLICANT: Kang, Sammo
; APPLICANT: Reinhard, Christoph
; APPLICANT: Jefferson, Anne Bennett
; TITLE OF INVENTION: POLYNUCLEOTIDES RELATED TO COLON CANCER
; FILE REFERENCE: 2300-1663
; CURRENT APPLICATION NUMBER: US/09/883.152
; CURRENT FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: 60/211.835
; PRIOR FILING DATE: 2000-06-15
; NUMBER OF SEQ ID NOS: 127
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 158
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-883-152-4

Query Match          58.1%; Score 36; DB 9; Length 158;
Best Local Similarity 41.7%; Pred. No. 29;
Matches 5; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
Db 97 GQTAEADLAWSCG 108

RESULT 12
US-09-747-835A-56
; Sequence 56, Application US/09747835A
; Patent No. US2002014692A1
; GENERAL INFORMATION:
; APPLICANT: Yamazaki, Victoria
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Zhou, Ping
; APPLICANT: Wang, Duncui
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyun
; APPLICANT: Asundi, Vinod
; APPLICANT: Drmanac, Radoje T
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO G PROTEIN-COUPLED RECEPTOR-LIKE
; TITLE OF INVENTION: LIKE) POLYPEPTIDES AND POLYNUCLEOTIDES

```

```

; FILE REFERENCE: HYS-37CIP
; CURRENT APPLICATION NUMBER: US/09/747.835A
; CURRENT FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: US 09/729.739
; PRIOR FILING DATE: 2000-12-04
; PRIOR APPLICATION NUMBER: US 09/653.450
; PRIOR FILING DATE: 2000-08-31
; PRIOR APPLICATION NUMBER: US 09/620.312
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: US 09/598.042
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/552.317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/488.725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 166
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-747-835A-56

Query Match          58.1%; Score 36; DB 10; Length 166;
Best Local Similarity 83.3%; Pred. No. 31;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAWACG 12
Db 30 TSWACG 35

RESULT 13
US-09-886-055-485
; Sequence 485, Application US/09886055
; Patent No. US20020132273A1
; GENERAL INFORMATION:
; APPLICANT: STRYER, LUBERT
; APPLICANT: ZOZULYA, SERGEY
; TITLE OF INVENTION: RECEPTOR FINGERPRINTING, SENSORY PERCEPTION, AND
; FILE REFERENCE: 078003-0277150
; CURRENT APPLICATION NUMBER: US/09/886.055
; CURRENT FILING DATE: 2001-06-22
; PRIOR APPLICATION NUMBER: 60/213.812
; PRIOR FILING DATE: 2000-06-22
; NUMBER OF SEQ ID NOS: 522
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 485
; LENGTH: 310
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-886-055-485

Query Match          58.1%; Score 36; DB 10; Length 310;
Best Local Similarity 83.3%; Pred. No. 52;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAWACG 12
Db 146 TSWACG 151

RESULT 14
US-10-184-644-549
; Sequence 549, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.

```

; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE OF INVENTION: ACIDS ENCODING THE SAME
 ; FILE REFERENCE: P3430R1C227
 ; CURRENT APPLICATION NUMBER: US/10/184,644
 ; CURRENT FILING DATE: 2002-06-28
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 549
 ; LENGTH: 2213
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 ; US-10-184-644-549

Query Match 58.1%; Score 36; DB 9; Length 2213;
 Best Local Similarity 57.1%; Pred. No. 2.8e+02;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

Qy 1 GAXXXETAMACGA 14
 Db 1752 GAGTCATACACGA 1765

RESULT 15
 US-09-280-197-11
 ; Sequence 11, Application US/09280197
 ; Patent No. US20020142403A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Yu, Shukun
 ; APPLICANT: Bojsen, Kirsten
 ; APPLICANT: Kragh, Karsten
 ; APPLICANT: Bojko, Maja
 ; APPLICANT: Nielsen, John
 ; APPLICANT: Marcussen, Jan
 ; APPLICANT: Christensen, Tove
 ; TITLE OF INVENTION: USE OF '-1,4-GLUCAN LYASE FOR PREPARATION OF
 ; TITLE OF INVENTION: 1,5-D-ANHYDROFRUCTOSE
 ; NUMBER OF SEQUENCES: 39
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Knobbde, Martens, Olsson & Bear
 ; STREET: 620 Newport Center Drive 16th Floor
 ; CITY: Newport Beach
 ; STATE: CA
 ; COUNTRY: U.S.A.
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/280.197
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/633,719
 ; FILING DATE: July 8, 1996
 ; APPLICATION NUMBER: PCT/EP94/03397
 ; FILING DATE: OCT-15-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Altman, Daniel E
 ; REGISTRATION NUMBER: 34,115
 ; REFERENCE/DOCKET NUMBER: DYO05.001APC
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 714-760-0404
 ; TELEFAX: 714-760-9502
 ; TELEX:
 ; INFORMATION FOR SEQ ID NO: 11:

; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 202 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 43
 ; OTHER INFORMATION: /note="X is a misc. amino acid"
 ; FEATURE:
 ; NAME/KEY: Modified-site
 ; LOCATION: 176
 ; OTHER INFORMATION: /note="X is a misc. amino acid"
 ; US-09-280-197-11

Query Match 56.5%; Score 35; DB 10; Length 202;
 Best Local Similarity 71.4%; Pred. No. 53;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 ETAMACG 12
 Db 68 DTAMNCG 74

Search completed: March 25, 2003, 08:23:40
 Job time : 15.1515 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:20:22 : Search time 15.2727 Seconds
(without alignments)
88.123 Million cell updates/sec

Title: US-09-646-532B-2
Perfect score: 62
Sequence: 1 GAXXXETAMACGXA 14

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR_73:*
2: pir1:*
3: pir2:*
4: pir3:*
5: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|--------|----------------------|
| 1 | 46 | 74.2 | 258 | 2 | F97387 | probable transcrip |
| 2 | 46 | 74.2 | 290 | 2 | AG2605 | transcription regu |
| 3 | 38 | 61.3 | 741 | 2 | AG3271 | hypothetical prote |
| 4 | 37 | 59.7 | 218 | 2 | G87548 | transglycosylase, |
| 5 | 37 | 59.7 | 275 | 2 | E75548 | conserved hypothet |
| 6 | 37 | 59.7 | 334 | 2 | T34634 | probable hydrolase |
| 7 | 37 | 59.7 | 479 | 2 | S73770 | NADH oxidase nox - |
| 8 | 37 | 59.7 | 502 | 2 | P85715 | hypothetical prote |
| 9 | 37 | 59.7 | 670 | 2 | AG5819 | hypothetical prote |
| 10 | 37 | 59.7 | 707 | 2 | C90999 | probable terminalase |
| 11 | 36 | 58.1 | 157 | 2 | S58000 | probable olfactory |
| 12 | 36 | 58.1 | 1533 | 2 | AA6221 | abdominal segment |
| 13 | 35 | 56.5 | 237 | 2 | JC7608 | type II lectin-like |
| 14 | 35 | 56.5 | 662 | 1 | AA9882 | histidine decarbox |
| 15 | 35 | 56.5 | 668 | 2 | T31633 | hypothetical prote |
| 16 | 35 | 56.5 | 1130 | 2 | AA8843 | MHC class II trans |
| 17 | 35 | 56.5 | 3165 | 2 | SI5010 | hypothetical prote |
| 18 | 34 | 54.8 | 209 | 2 | JC7239 | peroxiredoxin V - |
| 19 | 34 | 54.8 | 227 | 2 | BA8536 | DNA topoisomerase |
| 20 | 34 | 54.8 | 232 | 2 | B75561 | conserved hypothet |
| 21 | 34 | 54.8 | 269 | 2 | A34727 | Suirt-4 protein - m |
| 22 | 34 | 54.8 | 294 | 2 | F83371 | hypothetical prote |
| 23 | 34 | 54.8 | 297 | 2 | AE2805 | permease (imported |
| 24 | 34 | 54.8 | 321 | 2 | AE7584 | hypothetical prote |
| 25 | 34 | 54.8 | 356 | 2 | A25918 | thrombomodulin - b |
| 26 | 34 | 54.8 | 362 | 2 | H81821 | tRNA (uracil-5)-m |
| 27 | 34 | 54.8 | 370 | 2 | A97235 | RCCL repeats prote |
| 28 | 34 | 54.8 | 447 | 2 | D89646 | protein ZK455.8 [1 |
| 29 | 34 | 54.8 | 467 | 2 | E70564 | hypothetical prote |

| | | | | | |
|----|----|------|------|---|--------|
| 30 | 34 | 54.8 | 474 | 2 | S65763 |
| 31 | 34 | 54.8 | 501 | 2 | S48120 |
| 32 | 34 | 54.8 | 508 | 2 | B81212 |
| 33 | 34 | 54.8 | 516 | 2 | E81789 |
| 34 | 34 | 54.8 | 640 | 2 | D70850 |
| 35 | 34 | 54.8 | 672 | 2 | A65024 |
| 36 | 34 | 54.8 | 672 | 2 | H91046 |
| 37 | 34 | 54.8 | 672 | 2 | D85891 |
| 38 | 34 | 54.8 | 794 | 2 | T27870 |
| 39 | 34 | 54.8 | 1612 | 2 | S59969 |
| 40 | 34 | 54.8 | 1626 | 2 | A39242 |
| 41 | 33 | 53.2 | 124 | 2 | S66696 |
| 42 | 33 | 53.2 | 163 | 2 | G87331 |
| 43 | 33 | 53.2 | 246 | 2 | B82764 |
| 44 | 33 | 53.2 | 261 | 2 | A95894 |
| 45 | 33 | 53.2 | 276 | 2 | E87682 |

ALIGNMENTS

RESULT 1
F97387
probable transcription regulator (PA2489) [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence-revision 30-Sep-2001 #extLchange 11-Jan-2002
C:Accession: F97387
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirillo, B.; Goldm
A.; Liu, F.; Mollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: F97387
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-258 <KUR>
A:Cross-references: GB:AE007669; PIDN:AKK6055.1; PID:g1515128; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_407
A:Map position: circular chromosome

Query Match 74.2%; Score 46; DB 2; Length 258;
Best local similarity 64.3%; Pred. No. 0.48;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GAXXXETAMACGXA 14
DB 219 GASLAETAFACGFA 232

RESULT 2
AG2605
transcription regulator, Arac family Atu0239 [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence-revision 11-Jan-2002 #extLchange 11-Jan-2002
C:Accession: AG2605
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moc
etage, G.; Gillet, W.; Grant, C.; Gentilner, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCJ
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kar
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AG2605
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-290 <KUR>
A:Cross-references: GB:AE008688; PIDN:AL41261.1; PID:g17738567; GSPDB:GN00186
C:Genetics:
A:Experimental source: strain C58 (Dupont)
A:Gene: Atu0239
A:Map position: circular chromosome

Query Match 74.2% Score 46; DB 2; Length 290;
 Best Local Similarity 64.3%; Pred. No. 0.53;
 Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 GAXXETAMACGXA 14
 |||||
 Db 251 GASIAETAFACGFA 264

RESULT 3
 A:Accession: A83271
 hypothetical protein PA2984 [imported] - Pseudomonas aeruginosa (strain PA01)
 C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: A83271
 R:Stover, C.K.; Pham, X.O.; Ertvin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bt
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
 .; Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: A83271
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-741 <STO>
 A:Cross-references: GB:AE004724; GB:AE004091; NID:g9949083; PIDN:AG06372.1; GSPDB:GN001
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA2984

Query Match 61.3% Score 38; DB 2; Length 741;
 Best Local Similarity 85.7%; Pred. No. 35;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AMACGXA 14
 |||||
 Db 53 AMACGSA 59

RESULT 4
 G87548
 transglycosylase, probable CC2416 [imported] - Caulobacter crescentus
 C:Species: Caulobacter crescentus
 C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
 C:Accession: G87548
 R:Mermin, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
 B.; Laub, M.T.; Deboy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
 n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
 Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
 A:Title: Complete Genome Sequence of Caulobacter crescentus.
 A:Reference number: A87249; MUID:21173698; PMID:11259647
 A:Accession: G87548
 A:Molecule type: DNA
 A:Status: preliminary
 A:Residues: 1-718 <STO>
 A:Cross-references: GB:AE005673; NID:g13423957; PIDN:AAK24387.1; GSPDB:GN00148
 C:Genetics:
 A:Gene: CC2416

Query Match 59.7% Score 37; DB 2; Length 218;
 Best Local Similarity 85.7%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AMACGXA 14
 |||||
 Db 22 AMACGVA 28

RESULT 5
 E75548
 conserved hypothetical protein - Deinococcus radiodurans (strain R1)
 C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
 C:Accession: E75548
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J
 .; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266
 A:Accession: E75548
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-275 <WHI>
 A:Cross-references: GB:AE001882; GB:AE000513; NID:g6457865; PIDN:AAF09788.1; PID:g645
 A:Experimental source: strain R1
 C:Genetics:
 A:Gene: DR0203
 A:Map position: 1

Query Match 59.7% Score 37; DB 2; Length 275;
 Best Local Similarity 85.7%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AMACGXA 14
 |||||
 Db 163 AMACGFA 169

RESULT 6
 T34634
 probable hydrolase - Streptomyces coelicolor
 C:Species: Streptomyces coelicolor
 C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
 C:Accession: T34634
 R:Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream,
 submitted to the EMBL Data Library, June 1999
 A:Reference number: Z21549
 A:Accession: T34634
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-334 <SAU>
 A:Cross-references: EMBL:AL078618; PIDN:CAB44541.1; GSPDB:GN00070; SCOEDB:SC10A7.31
 A:Experimental source: strain A3(2)
 C:Genetics:
 A:Gene: SCOEDB:SC10A7.31

Query Match 59.7% Score 37; DB 2; Length 334;
 Best Local Similarity 57.1%; Pred. No. 26;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 GAXXETAMACGXA 14
 |||||
 Db 115 GASGTLEAACGSA 128

RESULT 7
 S73770
 NADH oxidase nox - Mycoplasma pneumoniae (strain ATCC 29342)
 N:Alternate names: hypothetical protein F11_orf479
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C:Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999
 C:Accession: S73770
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumo
 A:Reference number: S73327; MUID:97105885; PMID:8948633
 A:Accession: S73770
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-479 <HIM>
 A:Cross-references: EMBL:AE000044; GB:U00089; NID:g1674130; PIDN:AB96092.1; PID:g167
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:

A:Genetic code: SGC3
C:Superfamily: NADH peroxidase; dihydrolipoamide dehydrogenase homology

Query Match 59.7%; Score 37; DB 2; Length 479;
Best Local Similarity 50.0%; Pred. No. 36;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
| | | | |
DB 179 GLELAEMAMOCG 190

RESULT 8
F85715
hypothetical protein Z2116 [imported] - Escherichia coli (strain O157:H7, substrain EDLg
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: F85715
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: F85715
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-502 <STO>
A:Cross-references: GB:AE005174; NID:g12515071; PIDN:AAG56186.1; GSPDB:GN00145; UWGP:Z21
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z2116

Query Match 59.7%; Score 37; DB 2; Length 502;
Best Local Similarity 42.9%; Pred. No. 38;
Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 14
| | | | |
DB 178 GRDITDTKWRGCGA 191

RESULT 9
A85819
hypothetical protein Z3099 [imported] - Escherichia coli (strain O157:H7, substrain EDLg
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: A85819
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85819
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-670 <STO>
A:Cross-references: GB:AE005174; NID:g12516117; PIDN:AAG57013.1; GSPDB:GN00145; UWGP:Z30
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3099

Query Match 59.7%; Score 37; DB 2; Length 670;
Best Local Similarity 42.9%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 14
| | | | |
DB 178 GRDITDTKWRGCGA 191

RESULT 10
C90999
probable terminase large subunit [imported] - Escherichia coli (strain O157:H7, substrain

C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: C90999
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and 9
A:Reference number: A99629; MUID:21156331; PMID:11258796
A:Accession: C90999
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-707 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836386.1; PID:g13362432; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECS2963

Query Match 59.7%; Score 37; DB 2; Length 707;
Best Local Similarity 42.9%; Pred. No. 51;
Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 14
| | | | |
DB 215 GRDITDTKWRGCGA 228

RESULT 11
S58000
probable olfactory receptor tpcr71 - dog (fragment)
C:Species: Canis lupus familiaris (dog)
C>Date: 13-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 03-Nov-2000
C:Accession: S58000
R:Vanderhaeghen, P.; Schürmann, S.; Vaasatt, G.; Parmentier, M.
submitted to the EMBL Data Library, July 1995
A:Description: Male germ cells from several mammalian species express a specific repe
A:Reference number: S57995
A:Accession: S58000
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-157 <VAN>
A:Cross-references: EMBL:X89663; NID:g902194; PIDN:CA61810.1; PID:g902195
C:Superfamily: olfactory receptor OR14

Query Match 58.1%; Score 36; DB 2; Length 157;
Best Local Similarity 83.3%; Pred. No. 20;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 TAWACG 12
| | | | |
DB 22 TSMACG 27

RESULT 12
A46221
abdominal segment formation protein puntillo - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C>Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Oct-1997
C:Accession: A46221; S22026
R:Barker, D.D.; Wang, C.; Moore, J.; Dickinson, L.K.; Lehmann, R.
Genes Dev. 6, 2312-2326, 1992
A:Title: Puntillo is essential for function but not for distribution of the Drosophila
A:Reference number: A46221; MUID:93093466; PMID:1459455
A:Contents: emryo
A:Accession: A46221
A:Molecule type: mRNA
A:Status: preliminary
A:Residues: 1-153 <BAR>
A:Cross-references: GB:L07943; NID:g158190; PID:g158191
A:Note: sequence extracted from NCBI Backbone (NCBIN:120203, NCBI:120204)
R:Macdonald, P.M.
submitted to the EMBL Data Library, October 1991
A:Reference number: S22026
A:Accession: S22026

A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-361, 'A', 363-1102, 'R', 1104-1405, 'KN', 1408-1495, 'V', 1497-1518, 'S', 1520-1533
 A:Cross-references: EMBL:X62589; NID:g8393; PID:g8394
 C:Genetics:
 A:Gene: FlyBase:pum
 A:Cross-references: FlyBase:FBgn0003165

Query Match 58.1%; Score 36; DB 2; Length 1533;
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 ETAMAC 12
 Db 98 QTRWACG 104

RESULT 13
 JC7608
 Type II lectin-like immunoreceptor - human
 C:Species: Homo sapiens (man)
 C>Date: 30-Jun-2001 #sequence_revision 30-Jun-2001 #text_change 30-Jun-2001
 C:Accession: JC7608
 R:Hang, X.; Yuan, Z.; Chen, G.; Zhang, M.; Zhang, W.; Yu, Y.; Cao, X.
 Biochem. Biophys. Res. Commun. 281, 131-140, 2001
 A:Title: Cloning and characterization of a novel ITIM containing lectin-like immunoreceptor
 A:Reference number: JC7608; MUID:21092797; PMID:11178971
 A:Contents: Dendritic cells
 A:Accession: JC7608
 A:Molecule type: mRNA
 A:Residues: 1-237 <HDA>
 A:Cross-references: GB:AF067800
 C:Comment: This receptor, highly homologous to the group of macrophage/hepatic lectins 1 cell, especially in migrating, antigen capturing and processing.
 C:Genetics:
 A:Gene: 11lr
 A:Map position: 12p13
 F:45-65/Domain: transmembrane #status predicted <TM>

Query Match 56.5%; Score 35; DB 2; Length 237;
 Best Local Similarity 83.3%; Pred. No. 44;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 6 ETAMAC 11
 Db 100 ETAMSC 105

RESULT 14
 A49882
 histidine decarboxylase (EC 4.1.1.22) - human
 C:Species: Homo sapiens (man)
 C>Date: 30-Jun-1995 #sequence_revision 30-Jun-1995 #text_change 16-Jun-2000
 C:Accession: A49882; S11492; A56625
 R:Yatsunami, K.; Ohtsu, H.; Tsuchikawa, M.; Higuchi, T.; Ishibashi, K.; Shida, A.; Shima, J. Biol. Chem. 269, 1554-1559, 1994
 A:Title: Structure of the L-histidine decarboxylase gene.
 A:Reference number: A49882; MUID:94117478; PMID:8288622
 A:Accession: A49882
 A:Molecule type: DNA
 A:Residues: 1-662 <YAT>
 A:Cross-references: GB:D16583; NID:g516770; PIDN:BA04015.1; PID:g516771
 A:Note: only Intron-exon junctions shown
 R:Yamauchi, K.; Sato, R.; Tanno, Y.; Ohkawara, Y.; Maeyama, K.; Watanabe, T.; Sato, K.; Nucleic Acids Res. 18, 5891, 1990
 A:Title: Nucleotide sequence of the cDNA encoding L-histidine decarboxylase derived from
 A:Reference number: S11492; MUID:91016941; PMID:2216786
 A:Accession: S11492
 A:Molecule type: mRNA
 A:Residues: 1-147, 'Q', 149-662 <YAM>
 A:Cross-references: EMBL:X54297; NID:g32108; PIDN:CAA38196.1; PID:g32109
 R:Mamune-Sato, R.; Yamauchi, K.; Tanno, Y.; Ohkawara, Y.; Ohtsu, H.; Katayose, D.; Maeyama, J. Biochem. 209, 533-539, 1992

A:Title: Functional analysis of alternatively spliced transcripts of the human histid
 A:Reference number: S29263; MUID:93049295; PMID:1425659
 A:Accession: S29263
 A:Molecule type: mRNA
 A:Residues: 1-117, 'W', 119-147, 'Q', 149-499, 'W', 501-662 <MAN>
 R:Zahnov, C.A.; Yi, H.F.; McBride, O.W.; Joseph, D.R.
 DNA Seq. 1, 395-400, 1991
 A:Title: Cloning of the cDNA encoding human histidine decarboxylase from an erythrole
 A:Reference number: A56625; MUID:92119328; PMID:1768863
 A:Accession: A56625
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-662 <ZAH>
 A:Cross-references: GB:M60445; NID:g183924; PIDN:AAC41698.1; PID:g183925
 A:Experimental source: erythroleukemia cell line
 A:Note: sequence extracted from NCBI backbone (NCBIN:77739, NCBIPI:77742)
 C:Genetics:
 A:Gene: GDB:HDC
 A:Cross-references: GDB:128639; OMIM:142704
 A:Map position: 15pter-15qter
 A:Introns: 11/1; 68/3; 106/3; 147/3; 192/3; 240/3; 263/1; 317/2; 347/3; 380/3; 414/3
 C:Superfamily: human histidine decarboxylase; animal histidine decarboxylase homology
 C:Keywords: alternative splicing; carbon-carbon lyase; carboxy-lyase; phosphoprotein;
 F:3-471/Domain: animal histidine decarboxylase homology <HDC>
 F:15/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 56.5%; Score 35; DB 1; Length 662;
 Best Local Similarity 71.4%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 8 AMACGA 14
 Db 499 AMACGTS 505

RESULT 15
 T31633
 hypothetical protein Y57A10A.k - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
 C:Accession: T31633
 R:Smay, R.
 submitted to the EMBL Data Library, September 1999
 A:Reference number: Z21048
 A:Accession: T31633
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-668 <WTL>
 A:Cross-references: EMBL:AL117195; PIDN:CA855016.1; CESP:Y57A10A.k
 A:Experimental source: clone Y57A10A
 C:Genetics:
 A:Gene: CESP:Y57A10A.k
 A:Introns: 22/3; 182/3; 219/2; 340/3; 385/1; 451/3
 C:Superfamily: Caenorhabditis elegans hypothetical protein Y57A10A.k

Query Match 56.5%; Score 35; DB 2; Length 668;
 Best Local Similarity 45.5%; Pred. No. 1.1e+02;
 Matches 5; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 GAXXXETAMAC 11
 Db 326 GSTAHTGAC 336

Search completed: March 25, 2003, 08:22:19
 Job time : 18.2727 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:18:08 ; Search time 8.48485 Seconds
(without alignments)
68.436 Million cell updates/sec

Title: US-09-646-532B-2
Perfect score: 62
Sequence: 1 GAXXXETAMACGXA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|-------------|
| 1 | 37 | 59.7 | 479 | 1 | NAOX_MYCPN |
| 2 | 36 | 58.1 | 1533 | 1 | PUM_DROME |
| 3 | 35 | 56.5 | 312 | 1 | OAC1_HUMAN |
| 4 | 35 | 56.5 | 662 | 1 | DCHS_HUMAN |
| 5 | 35 | 56.5 | 1130 | 1 | C27A_HUMAN |
| 6 | 34 | 54.8 | 210 | 1 | POXS_MOUSE |
| 7 | 34 | 54.8 | 269 | 1 | SUR4_MOUSE |
| 8 | 34 | 54.8 | 356 | 1 | TRBM_BOVIN |
| 9 | 34 | 54.8 | 362 | 1 | TRMA_NEIMA |
| 10 | 34 | 54.8 | 404 | 1 | 118C_PIG |
| 11 | 34 | 54.8 | 501 | 1 | PHR1_SIGNAL |
| 12 | 34 | 54.8 | 640 | 1 | YOS1_MYCTU |
| 13 | 34 | 54.8 | 672 | 1 | HYRB_ECOLI |
| 14 | 34 | 54.8 | 1107 | 1 | ALIA2_ARATH |
| 15 | 34 | 54.8 | 1612 | 1 | TP2B_MOUSE |
| 16 | 34 | 54.8 | 1612 | 1 | TP2B_MOUSE |
| 17 | 34 | 54.8 | 1626 | 1 | TP2B_MOUSE |
| 18 | 34 | 54.8 | 1627 | 1 | TP2B_CHICK |
| 19 | 33 | 53.2 | 354 | 1 | TFE1_MACEU |
| 20 | 33 | 53.2 | 377 | 1 | C1Q1_XENLA |
| 21 | 33 | 53.2 | 404 | 1 | MOEA_HAEIN |
| 22 | 33 | 53.2 | 421 | 1 | KAS1_STRVN |
| 23 | 33 | 53.2 | 430 | 1 | SNAK_DROME |
| 24 | 33 | 53.2 | 493 | 1 | MOR1_MOUSE |
| 25 | 33 | 53.2 | 494 | 1 | MOR1_MOUSE |
| 26 | 33 | 53.2 | 494 | 1 | MOR1_MOUSE |
| 27 | 33 | 53.2 | 500 | 1 | MOR1_MOUSE |
| 28 | 33 | 53.2 | 521 | 1 | RRPB_IBVK |
| 29 | 33 | 53.2 | 545 | 1 | MUTL1_THETH |
| 30 | 33 | 53.2 | 583 | 1 | Y40C_RHISN |
| 31 | 33 | 53.2 | 598 | 1 | THIX_YEAST |
| 32 | 33 | 53.2 | 599 | 1 | THIX_YEAST |
| 33 | 33 | 53.2 | 604 | 1 | C1Q1_MOUSE |

| | | | | | | |
|----|----|------|------|---|------------|---------------------|
| 34 | 33 | 53.2 | 651 | 1 | BGLR_FELCA | O97524 felis alive |
| 35 | 33 | 53.2 | 669 | 1 | C1Q1_RAT | O920n7 rattus nov |
| 36 | 33 | 53.2 | 676 | 1 | C1Q1_HUMAN | P51787 homo sapien |
| 37 | 33 | 53.2 | 702 | 1 | EXO1_YEAST | P39875 saccharomyc |
| 38 | 33 | 53.2 | 2652 | 1 | RRPB_IBVK | P26314 avian infec |
| 39 | 33 | 53.2 | 2747 | 1 | FAP_DROME | P55824 drosophila |
| 40 | 32 | 51.6 | 242 | 1 | Y485_MYCTU | O53433 mycobacteri |
| 41 | 32 | 51.6 | 269 | 1 | SUR4_FUGRU | O57590 fungu rubrip |
| 42 | 32 | 51.6 | 382 | 1 | ISDF_CAUCR | O9a715 c lispd/lspf |
| 43 | 32 | 51.6 | 438 | 1 | SYN_THETH | P54263 thermus the |
| 44 | 32 | 51.6 | 1158 | 1 | ALAI_ARATH | P98204 arabidopsi |
| 45 | 32 | 51.6 | 2252 | 1 | POL1_GCMV | P13025 h rnal poly |

ALIGNMENTS

| RESULT 1 | ID | NAOX_MYCPN | STANDARD | PRT | 479 AA. |
|----------|--|------------|----------|-----|---------|
| AC | P75389 | | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | | |
| DT | 16-OCT-2001 (Rel. 40, Last annotation update) | | | | |
| DE | Probable NADH oxidase (EC 1.6.99.3) (NOXase). | | | | |
| GN | NOX OR MPN394 OR MP444. | | | | |
| OS | Mycoplasma pneumoniae. | | | | |
| OC | Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma. | | | | |
| OX | NCBI_Taxid=2104; | | | | |
| RN | (1) | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN-ATCC 29342 / M129; | | | | |
| RX | MEDLINE=97105885; PubMed=8948633; | | | | |
| RA | Hilbert R., Hilbert H., Plegens H., Plir E., Li B.-C., | | | | |
| RT | Hilbert R.; | | | | |
| RL | "Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae." | | | | |
| CC | Nucleic Acids Res. 24:4420-4449(1996). | | | | |
| CC | - FUNCTION: CATALYZES THE FOUR-ELECTRON REDUCTION OF MOLECULAR | | | | |
| CC | OXYGEN TO WATER (BY SIMILARITY). | | | | |
| CC | - CATALYTIC ACTIVITY: 2 NADH + O(2) -> 2 NAD(+) + 2 H(2)O. | | | | |
| CC | - COFACTOR: FAD. | | | | |
| CC | - SIMILARITY: SOME, TO FLAVOPROTEIN OXIDOREDUCTASES. HIGHEST, TO | | | | |
| CC | NADH PEROXIDASE (NPR). | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | | | |
| CC | use by non-profit institutions as long as its content is in no way | | | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/ | | | | |
| CC | or send an email to license@sib-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL: AEO00044; AAB96092.1; - | | | | |
| DR | HSSP: P37062; INHD. | | | | |
| DR | InterPro: IPR001327; FAD_Pyr_redox. | | | | |
| DR | InterPro: IPR000205; NAD_binding. | | | | |
| DR | InterPro: IPR004099; Pyr_redox_dim. | | | | |
| DR | PIfam: PF00070; Pyr_redox; 1. | | | | |
| DR | PIfam: PF02852; Pyr_redox_dim; 1. | | | | |
| DR | ProDom: PD000139; FAD_Pyr_redox; 1. | | | | |
| KW | Oxidoreductase; NAD; Flavoprotein; FAD; Redox-active center; | | | | |
| KW | Complete proteome. | | | | |
| FT | NP_BIND 3 33 | | | | |
| FT | ACT_SITE 11 11 | | | | |
| FT | ACT_SITE 43 43 | | | | |
| FT | NP_BIND 170 185 | | | | |
| FT | NP_BIND 295 305 | | | | |
| SO | SEQUENCE 479 AA; 52875 MW; 66886EA3BA8E5F1 CRC64; | | | | |

Query Match 59.7%; Score 37; DB 1; Length 479;
Best local Similarity 50.0%; Pred. No. 13;
Matches 6; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 GAXXXETAMACG 12
 Db 179 GUELEAAMQCG 190

RESULT 2

PUM_DROME STANDARD: PRT: 1533 AA.
 AC P25822:
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Maternal pumilio protein.
 GN PUM.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92249205; PubMed=1576962;
 RA Macdonald P.M.;
 RT "The Drosophila pumilio gene: an unusually long transcription unit
 RL and an unusual protein."
 RL Development 114:221-234(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93093466; PubMed=1459455;
 RA Barker D.D., Wang C., Moore J., Dickinson L.K., Lehmann R.;
 RT "Pumilio is essential for function but not for distribution of the
 RL Drosophila abdominal determinant Nanos."
 RL Genes Dev. 6:2312-2326(1992).
 CC -1- FUNCTION: PUM IS THE ONLY GENE REQUIRED FOR NOS ACTIVITY THAT IS
 CC NOT ALSO REQUIRED FOR POSTERIOR LOCALIZATION OF GERM LINE
 CC DETERMINANTS. PUM IS REQUIRED DURING EMERGOGENESIS WHEN NOS
 CC ACTIVITY APPARENTLY MOVES ANTERIORLY FROM THE POSTERIOR POLE.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. IT IS CONCENTRATED IN THE
 CC CORTICAL REGION OF THE EMBRYO BENEATH THE NUCLEI.
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE OVARIES AND DURING THE
 CC EMBRYOGENESIS.
 CC -1- DOMAIN: CONSISTS MAINLY OF REGIONS ENRICHED IN A SINGLE AMINO
 CC ACID.
 CC -1- DISEASE: LETHAL DEFECTIVE IN POSTERIOR PATTERN FORMATION.
 CC -1- SIMILARITY: BELONGS TO THE PUMILIO/MPIS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: X62589; CAA44474.1; -
 DR EMBL: L07943; AAB59189.1; -
 DR PIR: S22026; S22026.
 DR FlyBase: FBgn0003165; pum.
 DR InterPro: IPR001313; Pumilio/Puf.
 DR Pfam: PF00806; PUF.8.
 DR SMART: SM00025; Pumilio; 8.
 KW Developmental protein; Repeat.
 FT DOMAIN 34 45 ALA-RICH.
 FT DOMAIN 57 77 GLY/VAL-RICH.
 FT DOMAIN 83 93 ALA-RICH.
 FT DOMAIN 130 174 GLY-RICH.
 FT DOMAIN 152 164 POLY-GLY.
 FT DOMAIN 181 212 ALA-RICH.
 FT DOMAIN 213 236 GLN-RICH.
 FT DOMAIN 262 274 POLY-GLN.
 FT DOMAIN 571 599 GLY-RICH.
 FT DOMAIN 708 725 POLY-GLN.

FT DOMAIN 936 946 POLY-ALA.
 FT DOMAIN 1050 1062 POLY-ALA.
 FT DOMAIN 1111 1326 6 X 36 AA APPROXIMATE TANDEM REPEATS.
 FT REPEAT 1111 1146 1.
 FT REPEAT 1147 1182 2.
 FT REPEAT 1183 1218 3.
 FT REPEAT 1219 1254 4.
 FT REPEAT 1255 1290 5.
 FT REPEAT 1291 1326 6.
 FT CONFLICT 362 362 A -> S (IN REF. 2).
 FT CONFLICT 1103 1103 R -> P (IN REF. 2).
 FT CONFLICT 1406 1407 KN -> PH (IN REF. 2).
 FT CONFLICT 1496 1496 V -> I (IN REF. 2).
 FT CONFLICT 1519 1519 S -> G (IN REF. 2).
 SQ SEQUENCE 1533 AA; 157521 MW; C453A2321B8BDBC CRC64;

Query Match 58.1%; Score 36; DB 1; Length 1533;
 Best Local Similarity 71.4%; Pred. No. 57;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 6 ETAMACG 12
 Db 98 QTRMACG 104

RESULT 3

OAC1_HUMAN STANDARD: PRT: 312 AA.
 AC 096KK4:
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Olfactory receptor 10C1 (Hs6M1-17).
 GN OR10C1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Whitaker H.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases:
 CC -1- FUNCTION: PUTATIVE ODORANT RECEPTOR.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AL035542; CAB44507.1; ALT_INIT.
 DR InterPro: IPR000276; GPCR_Rhodpsn.
 DR Pfam: PF00001; 7tm_1; 1.
 DR PROSITE: PS00237; G_PROTEIN_RECPT_F1_1; 1.
 DR PROSITE: PS50262; G_PROTEIN_RECPT_F1_2; 1.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Multigene family; Olfaction.
 FT DOMAIN 1 24 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 25 48 1 (POTENTIAL).
 FT DOMAIN 49 56 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 57 78 2 (POTENTIAL).
 FT DOMAIN 79 99 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 100 119 3 (POTENTIAL).
 FT DOMAIN 120 138 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 139 157 4 (POTENTIAL).
 FT DOMAIN 158 194 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 195 218 5 (POTENTIAL).
 FT DOMAIN 219 235 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 236 258 6 (POTENTIAL).

```

FT DOMAIN 259 271 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 272 291 7 (POTENTIAL).
FT DOMAIN 292 312 CYTOPLASMIC (POTENTIAL).
FT DISULFID 96 188 BY SIMILARITY.
FT CARBOHYD 4 4 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 312 AA; 34351 MW; B6FD6E0E700CB2CC CRC64;

Query Match
Best Local Similarity 83.3%; Score 35; DB 1; Length 312;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAWACG 12
DB 146 SAWACG 151

RESULT 4
DCHS_HUMAN STANDARD; PRT; 662 AA.
ID P19113;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Histidine decarboxylase (EC 4.1.1.22) (HDC).
GN HDC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91016941; PubMed=2216786;
RA Yamuchi K., Ruziko S., Ohkawara Y., Tanno Y., Maeyama K.,
RA Matunabe T., Satoh K., Yoshizawa M., Shibahara S., Takishima T.;
RT "Nucleotide sequence of the cDNA encoding L-histidine decarboxylase
RL derived from human basophilic leukemia cell line, KU-812-F.";
RN Nucleic Acids Res. 18:5891-5891(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119328; PubMed=1768863;
RA Zhanow C.A., Yi H.F., McBride O.W., Joseph D.R.;
RT "Cloning of the cDNA encoding human histidine decarboxylase from an
RT erythroleukemia cell line and mapping of the gene locus to chromosome
RT 15.";
RN DNA Seq. 1:395-400(1991).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94117478; PubMed=8288622;
RA Yatsunami K., Ohtsu H., Tsuchikawa M., Higuchi T., Ishibashi K.,
RA Shida A., Shima Y., Nakagawa S., Yamauchi K., Yamamoto M.,
RA Hayashi N., Matunabe T., Ichikawa A.;
RT "Structure of the L-histidine decarboxylase gene.";
RN J. Biol. Chem. 269:1554-1559(1994).
CC -1- CATALYTIC ACTIVITY: L-histidine - histamine + CO(2).
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- PATHWAY: Histamine biosynthesis.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO GROUP II DECARBOXYLASES (DDC, GAD, HDC AND
CC TYRDC).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54297; CAA38196.1; -
CC DR EMBL: M60445; AAC41698.1; -
CC DR EMBL: D16583; BAA04015.1; -
CC DR PIR: S11492; S11492.
CC DR Genew: HGNC:4855; HDC.

```

```

DR MIM: 142704; -
DR InterPro: IPR002129; Pyridoxal dec.
DR Pfam: PF00282; Pyridoxal dec. 1.
DR PRINTS: PR00800; YHDCRBOXLASE.
DR PROSITE: PS00392; DDC_GAD_HDC_YDC; 1.
KW Lyase; Decarboxylase; Catecholamine biosynthesis; Pyridoxal phosphate.
FT BINDING 305 305 PYRIDOXAL PHOSPHATE (POTENTIAL).
FT CONFLICT 148 148 S -> Q (TN REF. 1)
SQ SEQUENCE 662 AA; 74140 MW; D7611CFAD60F469 CRC64;

Query Match
Best Local Similarity 71.4%; Score 35; DB 1; Length 662;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 8 AMACGX 14
DB 499 AMACGTS 505

RESULT 5
C27A_HUMAN STANDARD; PRT; 1130 AA.
ID C27A_HUMAN
AC P33076;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE MHC class II transactivator CIITA.
GN MHC27A OR CIITA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. AND VARIANTS.
RX MEDLINE=94006536; PubMed=8402893;
RA Steimle V., Otten L.A., Zufferey M., Mach B.;
RT "Complementation cloning of an MHC class II transactivator mutated in
RT hereditary MHC class II deficiency (or bare lymphocyte syndrome).";
RN Cell 75:135-146(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95269136; PubMed=7749984;
RA Riley J.L., Westerheide S.D., Price J.A., Brown J.A., Boss J.M.;
RT "Activation of class II MHC genes requires both the X box region and
RT the class II transactivator (CIITA).";
RN Immunity 2:533-543(1995).
CC -1- FUNCTION: ESSENTIAL FOR TRANSCRIPTIONAL ACTIVITY OF THE HLA CLASS
CC II PROMOTER. ACTIVATION IS VIA THE PROXIMAL PROMOTER. NO DNA
CC BINDING OF IN VITRO TRANSLATED CIITA WAS DETECTED. MAY ACT IN A
CC COACTIVATOR-LIKE FASHION THROUGH PROTEIN-PROTEIN INTERACTIONS BY
CC CONTACTING FACTORS BINDING TO THE PROXIMAL MHC CLASS II PROMOTER,
CC TO ELEMENTS OF THE TRANSCRIPTION MACHINERY, OR BOTH. ALTERNATIVELY
CC IT MAY ACTIVATE HLA CLASS II PROMOTER.
CC THAT BIND TO THE MHC CLASS II PROMOTER.
CC -1- DISEASE: DEFECTS IN MHC27A ARE A CAUSE OF HEREDITARY MHC CLASS II
CC DEFICIENCY (ALSO KNOWN AS BARE LYMPHOCYTE SYNDROME (BLS) OR HLA
CC CLASS II DEFICIENT COMBINED IMMUNODEFICIENCY). A FORM OF SEVERE
CC COMBINED IMMUNODEFICIENCY DISEASE (SCID). MHC27A IS LINKED WITH
CC BLS COMPLEMENTATION GROUP A.
CC -1- SIMILARITY: CONTAINS 1 NACHT DOMAIN.
CC -1- DATABASE: NACHT-Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/MHC27A1D260.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X74301; CAA52354.1; -
CC DR EMBL: U18259; AAA88861.1; -

```

```

DR Genew; HGNC:7067; MHC2TA.
DR MIM; 600005;
RT RT
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR Pfam; PF00560; LRR; 1.
DR SMART; SM00370; LRR; 4.
DR PROSITE; PS50837; NACHT; 1.
KW Transcription regulation; Activator; ATP-binding; Disease mutation;
KW SCID.
FT DOMAIN 414 724 NACHT.
FT 52 137 ASP/GLU-RICH (ACIDIC).
FT NP_BIND 420 427 ATP (POTENTIAL).
FT VARIANT 120 120 K -> IE (IN BLS).
FT /FTID=VAR_005127.
FT VARIANT 500 500 A -> G (IN BLS).
FT /FTID=VAR_005128.
FT VARIANT 940 963 MISSING (IN BLS).
FT /FTID=VAR_005129.
SQ SEQUENCE 1130 AA; 123456 MW; DC4D081802987E06 CRC64;

Query Match 56.5%; Score 35; DB 1; Length 1130;
Best Local Similarity 54.5%; Pred. No. 66;
Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AXXXTAMACG 12
Db 430 AGAVSRAMACG 440

RESULT 6
PDX5_MOUSE
ID PDX5_MOUSE STANDARD; PRT; 210 AA.
AC P99029; O90275; O90X45;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Antioxidant enzyme 5, mitochondrial precursor (Prx-V) (Peroxisomal
DE antioxidant enzyme) (PLP) (Thioredoxin peroxidase PMP20) (Antioxidant
DE enzyme B166) (AOEB166) (Liver tissue 2D-page spot 2D-00141V).
GN PRX5 OR PROX5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RX SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA Zhou Y., Kok K.H., Chun A.C.S., Wong C.M., Wu H.W., Lin M.C.M.,
RA Fung P.C.W., Kung H.-F., Jin D.-Y.;
RT "Mouse peroxidoredoxin V is a thioredoxin peroxidase that inhibits
RT p53-induced apoptosis.";
RL Biochem. Biophys. Res. Commun. 268:921-927(2000).
RN [2]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=994454545; PubMed=10514471;
RA Yamashita H., Avraham S., Jiang S., London R., Van Veldhoven P.P.,
RA Subramani S., Rogers R.A., Avraham H.;
RT "Characterization of human and murine PMP20 peroxisomal proteins that
RT exhibit antioxidant activity in vitro.";
RL J. Biol. Chem. 274:29897-29904(1999).
RN [3]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RX MEDLINE=994454545; PubMed=10514471;
RA Yamashita H., Avraham S., Jiang S., London R., Van Veldhoven P.P.,
RA Subramani S., Rogers R.A., Avraham H.;
RT "Cloning and characterization of AOEB166, a novel mammalian
RT antioxidant enzyme of the peroxidoredoxin family.";
RL J. Biol. Chem. 274:30451-30458(1999).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=20218665; PubMed=10753630;

```

```

RA Lee T.H., Kim S.J., Kang S.W., Lee K.K., Rhee S.G., Yu D.Y.;
RT "Molecular cloning and characterization of the mouse Peroxiredoxin V
RT gene.";
RL Biochem. Biophys. Res. Commun. 270:356-362(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Kidney;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi T., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I.,
RA Saito T., Okazaki Y., Gojodori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boilelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Hoftmann M., Hume D.A., Kamliya M., Lee N.H.,
RA Gustincich S., Hill D., Hofmann M., Fujita M., Gariboldi M.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilting L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/Genbank/DBJ databases.
RN [7]
RP SEQUENCE OF 50-61.
RC TISSUE=Liver;
RA Sanchez J.-C., Rouge V., Frutiger S., Hughes G.J., Yan J.X.,
RA Hoogland C., Appel R.D., Binz P.-A., Hochstrasser D.F.,
RA Cowthorne M.;
RL Submitted (AUG-1998) to the SWISS-PROT data bank.
CC -!- FUNCTION: Reduces hydrogen peroxide and alkyl hydroperoxides with
CC reducing equivalents provided through the thioredoxin system.
CC Involved in intracellular redox signalling.
CC -!- SUBUNIT: Monomer.
CC -!- SUBCELLULAR LOCATION: Mitochondrial, peroxisomal and cytoplasmic.
CC -!- ALTERNATIVE PRODUCTS: A SINGLE NUCLEAR GENE PRODUCES BOTH FORMS
CC BY USE OF ALTERNATIVE INITIATION CODONS IN THE SAME READING FRAME.
CC -!- TISSUE SPECIFICITY: Widely expressed.
CC -!- MISCELLANEOUS: ON THE 2D-GEL THE DETERMINED PI OF THIS UNKNOWN
CC PROTEIN IS: 7.6, ITS MW IS: 14.6 kDa.
CC -!- SIMILARITY: BELONGS TO THE PEROXIREDOXIN 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF197951; AAF04855.1; -
DR EMBL; AF124994; AAF27532.1; -
DR EMBL; AF110733; AAG13450.1; -
DR EMBL; AF208730; AAF21016.1; -
DR EMBL; AF208729; AAF21016.1; JOINED.
DR EMBL; AK002383; BAB22058.1; -
DR EMBL; AK003352; BAB22720.1; -
DR EMBL; BC008174; AAH08174.1; -
DR HSSP; P30044; 1HD2.
DR MGD; MGI:1859821; Prdx6.
DR SWISS-2DPAGE; P99029; MOUSE.
DR InterPro: IPR000866; Ahpc-TSA.
DR Pfam; PF00578; Ahpc-TSA; 1.

```

KW Antioxidant; Peroxisome; Mitochondrion; Transit peptide;
 KM Alternative Initiation.
 FT TRANSIT 1 210 MITOCHONDRION (POTENTIAL).
 FT CHAIN 2 210 PEROXIREDOXIN 5, MITOCHONDRIAL
 FT CHAIN 50 210 ISOFORM.
 FT INIT_MET 49 49 PEROXIREDOXIN 5, CYTOPLASMIC AND
 FT SITE 208 210 PEROXISOMAL ISOFORM.
 FT DISULFID 96 200 FOR CYTOPLASMIC AND PEROXISOMAL ISOFORM.
 FT CONFLICT 55 55 MICROBODY TARGETING SIGNAL (BY
 FT CONFLICT 83 102 SIMILARITY).
 FT TA (IN REF. 4).
 SO SEQUENCE 210 AA; 21897 MW; E944104CC468BDD8 CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 210;
 Best Local Similarity 46.2%; Pred. No. 22;
 Matches 6; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 2 AXXXXXETAMACGXA 14
 DB 27 AGRKAGMECGGA 39

RESULT 7
 SUR4_MOUSE
 ID SUR4_MOUSE STANDARD; PRT; 269 AA.
 AC 064310;
 DT 01-NOV-1997 (Rel. 35; Created)
 DT 01-NOV-1997 (Rel. 35; Last sequence update)
 DT 15-JUL-1999 (Rel. 38; Last annotation update)
 DE Surflet locus protein 4.
 GN SURF4 OR SURF-4.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid-10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/c;
 RC MEDLINE-90136574; PubMed-2300057;
 RA Huxley C., Fried M.;
 RT "The mouse surflet locus contains a cluster of six genes associated
 RT with four CpG-rich islands in 32 kilobases of genomic DNA.";
 RL Mol. Cell. Biol. 10:605-614(1990).
 RN [2]
 RP SUBCELLULAR LOCATION:
 RX MEDLINE-95316062; PubMed-7540914;
 RA Reeves J.E., Fried M.;
 RT "The surf-4 gene encodes a novel 30 kDa integral membrane protein.";
 RL Mol. Membr. Biol. 12:201-208(1995).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Endoplasmic
 CC reticulum.
 CC -1- SIMILARITY: BELONGS TO THE SURF4 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M62606; AAA40155.1; -
 CC EMBL: M62601; AAA40155.1; JOINED.
 CC EMBL: M62602; AAA40155.1; JOINED.
 CC EMBL: M62603; AAA40155.1; JOINED.
 CC EMBL: M62605; AAA40155.1; JOINED.
 CC EMBL: M63114; AAA40156.1; -
 CC MGD: MGI:98445; Surf4.
 CC InterPro: IPR002995; Surf4.
 CC Pfam: PF02077; SURF4; 1.

DR Prodom: PD010195; Surf4; 1.
 DR PROSITE: PS01339; SURF4; 1.
 KW Transmembrane; Endoplasmic reticulum.
 FT TRANSMEM 65 85 POTENTIAL.
 FT TRANSMEM 92 112 POTENTIAL.
 FT TRANSMEM 179 199 POTENTIAL.
 FT TRANSMEM 203 223 POTENTIAL.
 FT TRANSMEM 242 262 POTENTIAL.
 FT SITE 266 267 ENDOPASMIC RETICULUM RETRIEVAL MOTIF
 FT (POTENTIAL).
 SO SEQUENCE 269 AA; 30381 MW; CCIJF5219400E52F8 CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 269;
 Best Local Similarity 57.1%; Pred. No. 26;
 Matches 4; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 6 ETAMACG 12
 DB 56 DTTWMSG 62

RESULT 8
 TRBM_BOVIN
 ID TRBM_BOVIN STANDARD; PRT; 356 AA.
 AC P06579;
 DT 01-JAN-1988 (Rel. 06; Created)
 DT 01-JAN-1988 (Rel. 06; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Thrombomodulin (Fetomodulin) (TM) (Fragment).
 GN THBD.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_Taxid-9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-87067408; PubMed-3024152;
 RA Jackman R.W., Beeler D.L., Vandewater L., Rosenberg R.D.;
 RT "Characterization of a thrombomodulin cDNA reveals structural
 RT similarity to the low density lipoprotein receptor.";
 RL Proc. Natl. Acad. Sci. U.S.A. 83:8834-8838(1986).
 CC -1- FUNCTION: THROMBOMODULIN IS A SPECIFIC ENDOTHELIAL CELL RECEPTOR
 CC THAT FORMS A 1:1 STOICHIOMETRIC COMPLEX WITH THROMBIN. THIS
 CC COMPLEX IS RESPONSIBLE FOR THE CONVERSION OF PROTEIN C TO THE
 CC ACTIVATED PROTEIN C (PROTEIN CA). ONCE EVOLVED, PROTEIN CA
 CC SCISSIONS THE ACTIVATED COPROFACTORS OF THE COAGULATION MECHANISM,
 CC FACTOR VA AND FACTOR VIII, AND THEREBY REDUCES THE AMOUNT OF
 CC THROMBIN GENERATED.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ENDOTHELIAL CELLS ARE UNIQUE IN SYNTHESIZING
 CC THROMBOMODULIN.
 CC -1- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M14657; AAA30765.1; -
 CC PIR: A25918; A25918.
 CC HSSP: P07204; 1TMR.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001881; EGF_Ca.
 CC Pfam: PF00008; EGF; 5.
 CC SMART: SM00179; EGF_CA; 1.
 CC SMART: SM00001; EGF_like; 3.
 CC PROSITE: PS00010; ASX_HYDROXYL; 2.
 CC PROSITE: PS00022; EGF_1; FALSE_NEG.

DR PROSITE; PS01186; EGF_2; 3.
 DR PROSITE; PS01187; EGF_CA; 2.
 KW Endothelial cell; Receptor; Blood coagulation; Repeat; Transmembrane;
 KW Glycoprotein; EGF-like domain.
 FT NOK_TER 1 1
 FT DOMAIN 1 296 EXTRACELLULAR (POTENTIAL).
 FT TRANSEM 297 320 POTENTIAL.
 FT DOMAIN 321 356 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 17 57 EGF-LIKE 1.
 FT DOMAIN 60 98 EGF-LIKE 2.
 FT DOMAIN 99 137 EGF-LIKE 3, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 139 179 EGF-LIKE 4.
 FT DOMAIN 178 214 EGF-LIKE 5.
 FT DOMAIN 215 254 EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
 FT DISULFID 21 32 BY SIMILARITY.
 FT DISULFID 28 41 BY SIMILARITY.
 FT DISULFID 43 56 BY SIMILARITY.
 FT DISULFID 64 72 BY SIMILARITY.
 FT DISULFID 68 82 BY SIMILARITY.
 FT DISULFID 84 97 BY SIMILARITY.
 FT DISULFID 103 114 BY SIMILARITY.
 FT DISULFID 110 123 BY SIMILARITY.
 FT DISULFID 125 136 BY SIMILARITY.
 FT DISULFID 143 152 BY SIMILARITY.
 FT DISULFID 148 162 BY SIMILARITY.
 FT DISULFID 164 178 BY SIMILARITY.
 FT DISULFID 182 191 BY SIMILARITY.
 FT DISULFID 187 199 BY SIMILARITY.
 FT DISULFID 201 213 BY SIMILARITY.
 FT DISULFID 219 228 BY SIMILARITY.
 FT DISULFID 224 237 BY SIMILARITY.
 FT DISULFID 239 253 BY SIMILARITY.
 FT CARBOHYD 271 271 O-LINKED (GLYCOSAMINOGLYCAN) (BY SIMILARITY).
 SQ SEQUENCE 356 AA; 37795 MW; 29B41F097ABE1093 CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 356;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 8 AWACG 12
 ID 11111
 DB 18 AWACG 22
 RESULT 9
 TRMA_NEIMA STANDARD; PRT: 362 AA.
 AC Q9UT82;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE tRNA (Uracil-5)-methyltransferase (EC 2.1.1.35) (tRNA(M-5-U54) - methyltransferase) (R0MFT).
 GN TRMA OR NMA1938.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Z2491 / Serogroup A / Serotype 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Pakhill J., Achman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagers K., Leather S., Moule S., Mungall K., Quail M.A., Rajadaram M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrrell B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.";
 RT Nature 404:502-506(2000).
 CC -1- FUNCTION: Catalyzes the formation of 5-methyl-uridine at position 54 (M-5-U54) in all tRNA (By similarity).

CC -1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + tRNA = S-adenosyl-L-homocysteine + tRNA containing thymine.
 CC -1- SIMILARITY: BELONGS TO THE RNA M5U METHYLTRANSFERASE FAMILY. TRMA SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch).
 CC -----
 CC EMBL; AL162757; CAB85158.1; -
 CC InterPro; IPR000051; SAM_Bind.
 CC InterPro; IPR001566; TrMA.
 DR PROSITE; PS01230; TRMA_1; 1.
 DR PROSITE; PS01231; TRMA_2; FALSE_NEG.
 KW Transferase; Methyltransferase; tRNA processing; Complete proteome.
 FT DOMAIN 208 214 S-ADENOSYLMETHIONINE BINDING (BY SIMILARITY).
 FT ACT SITE 318 318 BY SIMILARITY.
 SQ SEQUENCE 362 AA; 40933 MW; B670B25D6FC5172 CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 362;
 Best Local Similarity 83.3%; Pred. No. 36;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 OY 9 WACGXA 14
 ID 11111
 DB 194 WACGXA 199
 RESULT 10
 ID IIBC_PIG STANDARD; PRT: 404 AA.
 AC Q9N211;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Interleukin-1 beta converting precursor (IL-1BC) (EC 3.4.22.36) (IL-1 beta converting enzyme) (ICE) (Interleukin-1 beta converting enzyme) (P45) (Caspase-1) (CASP-1).
 GN CASP1 OR IL1BC.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
 OX NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20039779; PubMed=10574622;
 RA Muneta Y., Shimoi Y., Yokomizo Y., Mori Y.;
 RT "Molecular cloning of porcine interleukin-1-beta converting enzyme and differential gene expression of IL-1beta converting enzyme, IL-1beta, and IL-18 in porcine alveolar macrophages.";
 RT J. Interferon Cytokine Res. 19:1289-1296(1999).
 RL J. Interferon Cytokine Res. 19:1289-1296(1999).
 CC -1- FUNCTION: THIOL PROTEASE THAT CLEAVES IL-1 BETA BETWEEN AN ASP AND AN ALA, RELEASING THE MATURE CYTOKINE WHICH IS INVOLVED IN A VARIETY OF INFLAMMATORY PROCESSES (By similarity).
 CC -1- CATALYTIC ACTIVITY: Release of interleukin 1-beta by specific cleavage at 116-Asp-1-Ala-117 and 27-Asp-1-Gly-28 bonds in precursor. Also hydrolyzes the small-molecule substrate, Ac-Tyr-Val-Ala-Asp-1-NHec.
 CC -1- SUBUNIT: Tetramer that consists of two heterodimers of a 20 kDa (P20) AND A 10 kDa (P10) SUBUNITS. P20 CAN ALSO FORM A HETERODIMER WITH THE EPSILON ISOFORM WHICH THEN HAS AN INHIBITORY EFFECT (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- PTM: THE TWO SUBUNTS ARE DERIVED FROM THE PRECURSOR SEQUENCE BY AN AUTOCATALYTIC MECHANISM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C14.
 CC -1- SIMILARITY: CONTAINS 1 CARD DOMAIN.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AB027296; BAA89531.1; -
 DR HSR: P29466; IICE.
 DR MEROPS: C14.001; -
 DR InterPro: IPR001315; CARD.
 DR InterPro: IPR002398; ICE.
 DR InterPro: IPR001318; ICE_P10.
 DR InterPro: IPR001309; ICE_P20.
 DR Pfam: PF00619; CARD; 1.
 DR Pfam: PF00655; ICE_P10; 1.
 DR Pfam: PF00656; ICE_P20; 1.
 DR PRINTS: PR00376; ILIACENZYM.
 DR SMART: SM00114; CARD; 1.
 DR SMART: SM00115; CASC; 1.
 DR PROSITE: PS50209; CARD; 1.
 DR PROSITE: PS01122; CASPASE_CYS; 1.
 DR PROSITE: PS01121; CASPASE_HIS; 1.
 DR PROSITE: PS50207; CASPASE_P10; 1.
 DR PROSITE: PS50208; CASPASE_P20; 1.
 KM Hydrolase: Thiol protease; Zymogen.
 FT PROPEP 1 119 POTENTIAL.
 FT CHAIN 120 297 INTERLEUKIN-1 BETA CONVERTASE P20.
 FT PROPEP 298 316 POTENTIAL.
 FT CHAIN 317 404 INTERLEUKIN-1 BETA CONVERTASE P10.
 FT DOMAIN 1 91 CARD.
 FT ACT_SITE 237 237 BY SIMILARITY.
 FT ACT_SITE 285 285 BY SIMILARITY.
 SQ SEQUENCE 404 AA; 44881 MW; 437DC787E85FB449 CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 404;
 Best Local Similarity 71.4%; Pred. No. 40;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 6 ETWACG 12
 DB 359 EHWACG 365
 RESULT 11
 PHRL_SINAL STANDARD; PRT; 501 AA.
 ID PHRL_SINAL
 AC P40115;
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 37, Last annotation update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Deoxyribodipyrroline photolase (EC 4.1.99.3) (DNA photolase)
 DE (Photoreactivating enzyme).
 GN PHRL.
 OS Sinapis alba (White mustard) (Brassicaceae).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Sinapis.
 OX NCBI_TaxId=3728;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94073224; PubMed=8252071;
 RA Batschauer A.;
 RT "A plant gene for photolase: an enzyme catalyzing the repair of UV-
 RT light-induced DNA damage.";
 RL Plant J. 4:705-709(1993).
 CC -1- FUNCTION: THIS ENZYME CATALYZES THE LIGHT-DEPENDENT MONOMERIZATION
 CC (300-600 NM) OF CYCLOBUTYL PYRIMIDINE DIMERS (IN CIS-SYN
 CC CONFIGURATION), WHICH ARE FORMED BETWEEN ADJACENT BASES ON THE
 CC SAME DNA STRAND, UPON EXPOSURE TO ULTRAVIOLET RADIATION.
 CC -1- CATALYTIC ACTIVITY: Cyclobutadipyrroline (in DNA) -> 2 pyrimidine
 CC residues (in DNA).

CC -1- COFACTOR: CONTAINS 2 CHROMOPHORES: A REDUCED FLAVIN (FADH2) AND AN
 CC OXIDIZED 8-HYDROXY-5-DEAZAFLAVIN (F420). BOTH CHROMOPHORES ARE
 CC BOUND BY NON-COVALENT INTERACTIONS (PROBABLE).
 CC -1- INDUCTION: BY VISIBLE LIGHT.
 CC -1- SIMILARITY: BELONGS TO THE DNA PHOTOLYASE CLASS-1 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X72019; CAA50898.1; -
 DR InterPro: IPR000474; DNA_photolase.
 DR InterPro: IPR002081; DNA_photolase_1.
 DR InterPro: IPR005101; FAD_binding_7.
 DR Pfam: PF00875; DNA_photolase; 1.
 DR Pfam: PF03441; FAD_binding_7; 1.
 DR ProDom: PD004390; DNA_photolase; 1.
 DR PROSITE: PS00394; DNA_PHOTOLYASES_1_1; 1.
 DR PROSITE: PS00691; DNA_PHOTOLYASES_1_2; 1.
 KW Lyase; Chromophore; Flavoprotein; FAD; DNA repair; DNA-binding.
 SQ SEQUENCE 501 AA; 57021 MW; C6B25CE0A33890DB CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 501;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 6 ETWAC 11
 DB 181 ETWAC 186
 RESULT 12
 Y051_MYCTU STANDARD; PRT; 640 AA.
 ID Y051_MYCTU
 AC Q10880; O53628;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV0083.
 GN RV0083 OR MT0090 OR MT0251.01 OR MT030.27.
 OS Mycobacterium tuberculosis.
 CC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
 CC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxId=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekla F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
 RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Bishel W., Ueberlack T., Weidman J., Khouri H., Gill J., Mikula A.,
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."

```

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021428; CAAL6264.1; -.
DR EMBL; AE006920; AAK44315.1; -.
DR TIGR; MT0090; -.
DR TubercuList; RV0083; -.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
KW Hypothetical protein; Oxidoreductase; Transmembrane;
KW Complete proteome.
FT TRANSMEM 8 28 POTENTIAL.
FT TRANSMEM 52 72 POTENTIAL.
FT TRANSMEM 90 110 POTENTIAL.
FT TRANSMEM 136 156 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 277 297 POTENTIAL.
FT TRANSMEM 298 318 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 391 411 POTENTIAL.
FT TRANSMEM 446 466 POTENTIAL.
FT TRANSMEM 497 517 POTENTIAL.
FT TRANSMEM 619 639 POTENTIAL.
SQ SEQUENCE 640 AA; 65627 MW; 6254C289BD108A8 CRC64;

Query Match 54.8%; Score 34; DB 1; Length 640;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 9 WACGXA 14
   |||||
Db 533 WACGAA 538

RESULT 13
HYFB_ECOLI
ID HYFB_ECOLI STANDARD; PRT; 672 AA.
AC P23482;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrogenase-4 component B (EC 1.-.-.-).
GN HYFB OR B2482.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN-K12;
RA Andrews S.C., McClay J., Ambler A., Quail M., Berks B.C., Guest J.R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
```

```

RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakada S., Nakamura Y., Nashimoto H.,
RA Osima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tadami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yanagata S., Horiiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 30.0-68.8 min on the linkage map and
RT analysis of its sequence features."
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE OF 1-34 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=91202105; PubMed=2016588;
RA Andrews S.C., Harrison P.M., Guest J.R.;
RT "A molecular analysis of the 53.3 minute region of the Escherichia
RT coli linkage map."
RL J. Gen. Microbiol. 137:361-367(1991).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (by similarity).
CC -1- SIMILARITY: TO POLYPEPTIDE 5 OF THE NADH-UBIQUINOL OXIDOREDUCTASE
CC OF CHLOROPLASTS OR MITOCHONDRIA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M63654; AAB88564.1; -.
DR EMBL; AE000335; AAC75355.1; -.
DR EMBL; D90876; BAA16360.1; -.
DR EMBL; D90877; BAA16370.1; -.
DR PIR; D49749; D49749.
DR Ecogene; EG11282; hyfb.
DR InterPro: IPR001750; Oxidored_q1.
DR Pfam; PF00361; oxidored_q1; 1.
KW Oxidoreductase; Transmembrane; Inner membrane; Complete proteome.
FT TRANSMEM 6 26 POTENTIAL.
FT TRANSMEM 31 51 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 165 185 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 312 332 POTENTIAL.
FT TRANSMEM 343 363 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
FT TRANSMEM 475 495 POTENTIAL.
FT TRANSMEM 505 525 POTENTIAL.
FT TRANSMEM 532 552 POTENTIAL.
FT TRANSMEM 652 672 POTENTIAL.
SQ SEQUENCE 672 AA; 72582 MW; 3CA8C2BD25E54B1B CRC64;

Query Match 54.8%; Score 34; DB 1; Length 672;
Best Local Similarity 100.0%; Pred. No. 63;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 AWACG 12
   |||||
Db 565 AWACG 569

RESULT 14
```


ALA2_ARATH STANDARD: PRT: 1107 AA.
 ID ALA2_ARATH
 AC P98205:
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Potential phospholipid-transporting ATPase 2 (EC 3.6.3.1)
 DE (Amnophospholipid flippase 2).
 GN ALA2 OR AT5G44240 OR MNL1.17.
 -OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 OC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
 OC eurosids II: Brassicales: Brassicaceae: Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=97471969; PubMed=9330910;
 RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
 RA Miyajima N., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
 RT features of the 1.6 Mb regions covered by twenty physically assigned
 RT pl clones."
 RL DNA Res. 4:215-230(1997).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RA Axelsson K.B.;
 RL Unpublished observations (NOV-2000).
 CC -1- FUNCTION: INVOLVED IN TRANSPORT OF PHOSPHOLIPIDS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (E1-E2 ATPASES) SUBFAMILY IV
 CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO WRONG
 CC PREDICTIONS OF EXONS AND BOTH TERMINI FROM THE GENOMIC SEQUENCE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB005239; BAB10991.1; ALT_SEQ.
 DR InterPro: IPR001757; ATPase_E1-E2.
 DR InterPro: IPR001454; Hlgase/hydriase.
 DR Pfam: PF00702; Hydrolase; 1.
 DR PRINTS: PR00119; CATAATPASE.
 DR PROSITE: PS00154; ATPASE_E1_E2; 1.
 KW Hydrolase; Transmembrane; Phosphorylation; ATP-binding; Metal-binding;
 KW Magnesium; Multigene family.
 FT DOMAIN 1
 FT TRANSMEM 34 53
 FT DOMAIN 56 60
 FT TRANSMEM 61 83
 FT DOMAIN 84 268
 FT TRANSMEM 269 290
 FT DOMAIN 291 315
 FT TRANSMEM 316 333
 FT TRANSMEM 334 807
 FT TRANSMEM 808 827
 FT DOMAIN 828 841
 FT TRANSMEM 842 860
 FT TRANSMEM 861 890
 FT TRANSMEM 891 912
 FT TRANSMEM 913 919
 FT TRANSMEM 920 942
 FT TRANSMEM 943 948
 FT TRANSMEM 949 969
 FT DOMAIN 970 982
 FT TRANSMEM 983 1007
 FT DOMAIN 1008 1107

FT MOD_RES 381 381 PHOSPHORYLATION (BY SIMILARITY).
 FT METAL 752 752 MAGNESIUM (BY SIMILARITY).
 FT METAL 756 756
 SQ SEQUENCE 1107 AA; 124835 MW; 31E7729E9653C96F CRC64;
 Query Match 54.8%; Score 34; DB 1; Length 1107;
 Best local similarity 83.3%; Pred. No. 99;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 7 TAAACG 12
 Db 231 TEWACG 236
 RESULT 15
 TP2B_CRILLO STANDARD: PRT: 1612 AA.
 ID TP2B_CRILLO
 AC 064399;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA topoisomerase II, beta isozyme (EC 5.99.1.3).
 DE TOP2B.
 GN Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OS Cricetulus longicaudatus (Long-tailed hamster) (Chinese hamster).
 OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
 OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Cricetinae;
 OC Cricetus.
 OX NCBI_TaxID=10030;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=96085121; PubMed=7495861;
 RA Derendre S., Frey S., Delaporte C., Jacquemin-Sablon A.;
 RT "Cloning and characterization of full-length cDNAs coding for the DNA
 RT topoisomerase II beta from Chinese hamster lung cells sensitive and
 RT resistant 9-OH-ellipticine."
 RL Biochim. Biophys. Acta 1264:178-182(1995).
 CC -1- FUNCTION: CONTROL OF TOPOLOGICAL STATES OF DNA BY TRANSIENT
 CC BREAKAGE AND SUBSEQUENT REJOINING OF DNA STRANDS. TOPOISOMERASE II
 CC MAKES DOUBLE-STRAND BREAKS.
 CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining
 CC of double-stranded DNA.
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear; nucleolar.
 CC -1- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
 CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
 CC RELAX ONLY NEGATIVE SUPERCOILS.
 CC -1- SIMILARITY: BELONGS TO THE TYPE II TOPOISOMERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X86455; CA60173.1; -;
 DR HSSP: P06786; IBGM.
 DR InterPro: IPR003594; ATPbind_ATPase.
 DR InterPro: IPR003957; CBFA_NFYB_topis.
 DR InterPro: IPR001241; DNA_topoisoi.
 DR InterPro: IPR002205; DNA_topoisoi.
 DR Pfam: PF00204; DNA_gyraseb; 1.
 DR Pfam: PF00521; DNA_topoisoi; 1.
 DR Pfam: PF02518; HATPase_c; 1.
 DR PRINTS: PR00615; CCAATSUBUNTA.
 DR PRINTS: PR00418; TP12FAMILY
 DR PRODOM: PD000616; DNA_topoisoi; 1.
 DR PRODOM: PD000742; DNA_topoisoi; 1.
 DR SMART: SM00433; TOP2c; 1.
 DR SMART: SM00434; TOP4c; 1.
 DR PROSITE: PS00177; TOPOISOMERASE_II; 1.

KW Isomerase; Topoisomerase; DNA-binding; ATP-binding; Nuclear protein.
 NP_BIND 170 175 ATP (POTENTIAL).
 FT ACT_SITE 814 814 DNA CLEAVAGE (BY SIMILARITY).
 SQ SEQUENCE 1612 AA; 182074 MW; C01D6FC40620FC68 CRC64;

Query Match 54.8%; Score 34; DB 1; Length 1612;
 Best Local Similarity 54.5%; Pred. No. 1.4e+02;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 GAXXXETAWAC 11
 |||
 Db 861 GAGGICGTGWAC 871

Search completed: March 25, 2003, 08:20:12
 Job time : 10.4848 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:19:47 : Search time 30.9697 seconds
(without alignments)
93.145 Million cell updates/sec

Title: US-09-646-532b-2
Perfect score: 62
Sequence: 1 GAXXXETAMACGXA 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-----------|---------------------|
| 1 | 46 | 74.2 | 290 | 16 Q8U104 | O8U104 agrobacteri |
| 2 | 40 | 64.5 | 276 | 5 O97398 | O97398 phaeodon coc |
| 3 | 40 | 64.5 | 423 | 2 O9KHK6 | O9KHK6 streptomyc |
| 4 | 39 | 62.9 | 719 | 3 O96WJ0 | O96WJ0 pneumocysti |
| 5 | 39 | 62.9 | 1681 | 4 O9UGL1 | O9UGL1 homo sapien |
| 6 | 38 | 61.3 | 195 | 16 O92K85 | O92K85 rhizobium m |
| 7 | 38 | 61.3 | 229 | 9 O9XJY2 | O9XJY2 streptococ |
| 8 | 38 | 61.3 | 229 | 9 O9WCH2 | O9WCH2 streptococ |
| 9 | 38 | 61.3 | 741 | 16 O9H2M0 | O9H2M0 pseudomonas |
| 10 | 37 | 59.7 | 218 | 16 O9A5N1 | O9A5N1 caulobacter |
| 11 | 37 | 59.7 | 275 | 16 O9RXV3 | O9RXV3 delinococcus |
| 12 | 37 | 59.7 | 334 | 16 O9XAP4 | O9XAP4 streptomyc |
| 13 | 37 | 59.7 | 448 | 2 O9RH01 | O9RH01 lactobacilli |
| 14 | 37 | 59.7 | 502 | 16 O9XAB7 | O9XAB7 escherichia |
| 15 | 37 | 59.7 | 670 | 2 O8VNV7 | O8VNV7 escherichia |
| 16 | 37 | 59.7 | 707 | 16 O9EYD1 | O9EYD1 escherichia |

| | | | | | |
|----|----|------|------|-----------|--------------------|
| 17 | 36 | 58.1 | 157 | 6 O28305 | O28305 canis famli |
| 18 | 36 | 58.1 | 217 | 4 O43885 | O43885 homo sapien |
| 19 | 36 | 58.1 | 217 | 4 O43888 | O43888 homo sapien |
| 20 | 36 | 58.1 | 217 | 4 O96R48 | O96R48 homo sapien |
| 21 | 36 | 58.1 | 267 | 5 O9VHM4 | O9VHM4 drosophila |
| 22 | 36 | 58.1 | 336 | 2 O8VUZ3 | O8VUZ3 bradyrhizob |
| 23 | 35 | 56.5 | 113 | 5 O61052 | O61052 trypanosoma |
| 24 | 35 | 56.5 | 216 | 4 O96R18 | O96R18 homo sapien |
| 25 | 35 | 56.5 | 230 | 4 O8XW9 | O8XW9 homo sapien |
| 26 | 35 | 56.5 | 237 | 4 O9NS33 | O9NS33 homo sapien |
| 27 | 35 | 56.5 | 237 | 4 O9U134 | O9U134 homo sapien |
| 28 | 35 | 56.5 | 237 | 4 O9UMR7 | O9UMR7 homo sapien |
| 29 | 35 | 56.5 | 342 | 8 O9G650 | O9G650 otocryptis |
| 30 | 35 | 56.5 | 357 | 10 O8RUN0 | O8RUN0 oryza sativ |
| 31 | 35 | 56.5 | 417 | 2 O9F6E0 | O9F6E0 streptomyc |
| 32 | 35 | 56.5 | 423 | 2 O9XE00 | O9XE00 actinomdur |
| 33 | 35 | 56.5 | 424 | 16 O92S34 | O92S34 rhizobium m |
| 34 | 35 | 56.5 | 484 | 4 O8WWR8 | O8WWR8 homo sapien |
| 35 | 35 | 56.5 | 518 | 16 O8XZV3 | O8XZV3 ralsionia s |
| 36 | 35 | 56.5 | 630 | 10 O9FUI1 | O9FUI1 oryza sativ |
| 37 | 35 | 56.5 | 710 | 5 O9NA69 | O9NA69 caenorhabd |
| 38 | 35 | 56.5 | 804 | 4 O96SN3 | O96SN3 homo sapien |
| 39 | 35 | 56.5 | 884 | 7 O29675 | O29675 homo sapien |
| 40 | 35 | 56.5 | 928 | 2 O9ZG10 | O9ZG10 streptomyc |
| 41 | 35 | 56.5 | 932 | 4 O96K14 | O96K14 homo sapien |
| 42 | 35 | 56.5 | 1092 | 10 O9STC2 | O9STC2 gracilariop |
| 43 | 35 | 56.5 | 1141 | 12 O82456 | O82456 bayou virus |
| 44 | 35 | 56.5 | 1208 | 4 O9HCF8 | O9HCF8 homo sapien |
| 45 | 35 | 56.5 | 3164 | 12 O9YTU2 | O9YTU2 cryphonectr |

ALIGNMENTS

RESULT 1
ID O8U104 PRELIMINARY: PRT: 290 AA.
AC O8U104:
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DE Transcriptional regulator, Arac family.
GN ATU0239 OR AGR.C.407.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria: Proteobacteria: alpha subdivision: Rhizobiaceae group:
OC Rhizobiaceae: Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gilliet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McCelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Petry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.,
RA "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Guroilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houtmel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA Wollan C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Stear C., Strub G.,
RA Cielo C., Slater S.,
RT "Genome sequence of the plant pathogen and biotechnology agent

```

RT Agrobacterium tumefaciens C58.;
RL Science 294:2323-2328(2001).
DR EMBL: AE008996; ALA41261.1; -.
DR EMBL: AE007963; AKK86055.1; ALT_INIT.
KW Complete proteome.
SQ SEQUENCE 290 AA; 32633 MW; C6F3643FA611B58A CRC64;

Query Match
Best Local Similarity 64.3%; Score 46; DB 16; Length 290;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 14
DB 251 GASLAETAFACGFA 264

RESULT 2
ID 097398 PRELIMINARY; PRT; 276 AA.
AC 097398;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Chymotrypsin precursor.
OS Phaedon cochleariae (Mustard beetle).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Coleoptera; Polyphaga;
OC Cucujiformia; Phytophaga; Chrysomeloidea; Chrysomelidae;
OC Chrysomelinae; Phaedon.
OX NCBI_TaxID=80249;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=GUT;
RA Gizard C., Jouanin L.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY
DR EMBL: Y17904; CAI6928.1; -.
DR HSSP; P00761; LEPT.
DR MEROPS; S01.122; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Ser. protease_Try.
DR Pfam: PF00089; trypsin_1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPE; 1.
DR PROSITE; PS50240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Hydrolyase; Serine protease; Signal.
FT SIGNAL 1
FT CHAIN 46 276 CHYMOTRYPSIN.
SQ SEQUENCE 276 AA; 29868 MW; FC5FD05DB882A1DE CRC64;

Query Match
Best Local Similarity 54.5%; Score 40; DB 5; Length 276;
Matches 6; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 AXXXXETAMACG 12
DB 65 ASAGETSWTCG 75

RESULT 3
ID 09KH6 PRELIMINARY; PRT; 423 AA.
AC 09KH6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative keto synthase alpha Enca.
ENCA.
GN Streptomyces maritimus.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

```

```

OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=115828;
RN [1]
RP SEQUENCE FROM N.A.
RA Piel J., Hoang K., Moore B.S.;
RT "Natural Metabolic Diversity Encoded by the Enterocin Biosynthesis
Gene Cluster.";
RL J. Am. Chem. Soc. 122:5415-5416(2000).
DR EMBL; AF254925; AAF81728.1; -.
DR HSSP; P73283; IESM.
DR InterPro: IPR000794; Ketoacyl-synt.
DR Pfam; PF00109; ketoacyl-synt.1.
DR Pfam; PF02801; ketoacyl-synt.C; 1.
DR PROSITE; PS00606; B_KETOACYL_SYNTHASE; UNKNOWN_1.
SQ SEQUENCE 423 AA; 44640 MW; A998CC6B2501F7DA CRC64;

Query Match
Best Local Similarity 58.3%; Score 40; DB 2; Length 423;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 GAXXXETAMACG 12
DB 147 GSLARETAMACG 158

RESULT 4
ID 096WJ0 PRELIMINARY; PRT; 719 AA.
AC 096WJ0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Lanosterol synthase.
GN ERG7.
OS Pneumocystis carinii.
OC Eukaryota; Fungi; Ascomycota; Pneumocystidomycetes; Pneumocystidaceae.
OC Pneumocystis.
OX NCBI_TaxID=4754;
RN [1]
RP SEQUENCE FROM N.A.
RA Joubert B.M., Matsuda S.P.T.;
RT "Cloning and characterization of an oxidosqualene cyclase cDNA
establish Lanosterol as a Pneumocystis carinii metabolite.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF285825; AAK82993.1; -.
DR InterPro: IPR001330; Prenyltrans.
DR Pfam; PF00432; prenyltrans; 3.
SQ SEQUENCE 719 AA; 83308 MW; 4904DA1B8829F675 CRC64;

Query Match
Best Local Similarity 52.9%; Score 39; DB 3; Length 719;
Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 GAXXXETAMAC 11
DB 639 GSQVOTAMAC 649

RESULT 5
ID 09UGL1 PRELIMINARY; PRT; 1681 AA.
AC 09UGL1;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RB-binding protein.
GN RBBP2H1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.

```

RA Kashuba V., Protodopov A., Podowski R., Zabarovsky E.;
 RT "Isolation and chromosomal localization of a new human retinoblastoma
 binding protein 2 homolog 1a (RBBP2H1A)";
 RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ243706; CAB63108.1; -;
 DR InterPro: IPR001606; ARID.
 DR InterPro: IPR003347; TF_JmjC.
 DR InterPro: IPR003449; TF_JmjN.
 DR InterPro: IPR004198; znf_C5HC2.
 DR InterPro: IPR001965; znf_PHD.
 DR Pfam: PF01388; ARID.1.
 DR Pfam: PF02373; JmjC.1.
 DR Pfam: PF02375; JmjN.1.
 DR Pfam: PF00628; PHD.3.
 DR Pfam: PF02928; zf-C5HC2.1.
 DR SMART: SM00501; BRIGHT.1.
 DR SMART: SM00545; JmjN.1.
 DR SMART: SM00249; PHD.3.
 SQ SEQUENCE 1681 AA; 190118 MW; 0429FA1E7BAFB702 CRC64;

Query Match
 Best Local Similarity 62.9%; Score 39; DB 4; Length 1681;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 TAWACG 12
 |||||
 DB 77 TAWACG 82

RESULT 6
 ID 092K85 PRELIMINARY; PRT; 195 AA.
 AC 092K85;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical transmembrane protein SMC04211.
 GN R01881 OR SMC04211.
 OS Rhizobium meliloti (Sinorhizobium meliloti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Sinorhizobium.
 OX NCBI_TaxID=382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396507; PubMed-11481430;
 RA Capela D., Barloy-Hudler F., Gouzy J., Bothe G., Ampe F., Batut J.,
 RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
 RA Godie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
 RA Pohl T., Portetelle D., Puehler A., Purnelle B., Rampsberger U.,
 RA Renard C., Thebaud P., Vandenbol M., Weidner S., Gallibert F.;
 RT "Analysis of the chromosome sequence of the legume symbiont
 Sinorhizobium meliloti strain 1021.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
 DR EMBL: AL591788; CAC46460.1; -;
 DR InterPro: IPR000080; SNAse.
 DR Pfam: PF00565; SNAse.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 195 AA; 21051 MW; 687635673D5031FF CRC64;

Query Match
 Best Local Similarity 61.3%; Score 38; DB 16; Length 195;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 6 ETAWACG 12
 |||||
 DB 69 ETAWACG 75

RESULT 7
 ID 09XJV2 PRELIMINARY; PRT; 229 AA.
 AC 09XJV2;

DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE ORF229 gp.
 GN ORF229.
 OS Streptococcus thermophilus bacteriophage SF119.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=72638;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98160788; PubMed-9499809;
 RA Desiere F., Lucchini S., Brusow H.;
 RT "Evolution of Streptococcus thermophilus bacteriophage genomes by
 RT modular exchanges followed by point mutations and small deletions and
 RT insertions.";
 RL Virology 241:345-356(1998).
 DR EMBL: AF115102; AAD44067.1; -;
 SQ SEQUENCE 229 AA; 25702 MW; C6AC1CA562CDD37 CRC64;

Query Match
 Best Local Similarity 61.3%; Score 38; DB 9; Length 229;
 Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GAXXETAWACGXA 14
 |||||
 DB 118 GARVGNAMWCGNA 131

RESULT 8
 ID 09MCH2 PRELIMINARY; PRT; 229 AA.
 AC 09MCH2;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Gp229.
 OS Streptococcus thermophilus bacteriophage SF118.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
 OX NCBI_TaxID=74382;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99412383; PubMed-10482618;
 RA Lucchini S., Desiere F., Brusow H.;
 RT "Comparative genomics of Streptococcus thermophilus phage species
 RT supports a modular evolution theory.";
 RL J. Virol. 73:8647-8656(1999).
 DR EMBL: AF158601; AAF63082.1; -;
 SQ SEQUENCE 229 AA; 25682 MW; ED2AD9129827BD95 CRC64;

Query Match
 Best Local Similarity 61.3%; Score 38; DB 9; Length 229;
 Matches 7; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

OY 1 GAXXETAWACGXA 14
 |||||
 DB 118 GARVGNAMWCGNA 131

RESULT 9
 ID 09HZM0 PRELIMINARY; PRT; 741 AA.
 AC 09HZM0;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Hypothetical protein PA2984.
 GN PA2984.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OX Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE-20437337; PubMed-10984043;
 RA Stover C.R., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Gardner L.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laddig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RT "Complete genome sequence of *Pseudomonas aeruginosa* PA01, an
 opportunistic pathogen.";
 RL Nature 406:959-964(2000).
 DR EMBL, AE004724; AAC06372.1; -;
 DR InterPro: IPR001279; Blackmase-1like.
 DR InterPro: IPR004477; COMEC_N-term.
 DR InterPro: IPR004797; COMEC_N-term.
 DR Pfam: PF00753; lactamase.B; 1.
 DR TIGRFAMS: TIGR00360; COMEC_N-term; 1.
 DR TIGRFAMS: TIGR00361; COMEC_Recc2; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 741 AA; 79979 MW; 2c17d3850CEB98D0 CRC64;

Query Match 61.3%; Score 38; DB 16; Length 741;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AWACGXA 14
 |||||
 Db 53 AWACGSA 59

RESULT 10
 O9A5N1 PRELIMINARY: PRT; 218 AA.
 ID O9A5N1;
 AC O9A5N1;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Transglycosylase, putative.
 GN CC2416.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
 CC Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 19089 / CB15;
 RX MEDLINE-21173698; PubMed-11259647;
 RA Niernman W.C., Feldblyum T.V., laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA Deboy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
 RA Kolony J.F., Smit J., Craven M.B., Khouli H., Shetty J., Berry K.,
 RA Uterback T., Tran K., Wolf A., Yamatchyan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.,
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
 DR EMBL, AE005911; AAK24387.1; -;
 DR HSSP; P03810; 10SA.
 DR TIGR; CC2416; -;
 DR InterPro: IPR000189; SUL_domain.
 DR Pfam; PF01464; SUL; 1.
 DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.
 KW Complete proteome.
 SQ SEQUENCE 218 AA; 23059 MW; 7E3106D6FED6BB7 CRC64;

Query Match 59.7%; Score 37; DB 16; Length 218;
 Best Local Similarity 85.7%; Pred. No. 46;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AWACGXA 14
 |||||
 Db 22 AWACGSA 28

RESULT 11
 O9RXV3 PRELIMINARY: PRT; 275 AA.
 ID O9RXV3;
 AC O9RXV3;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein DR0203.
 GN DR0203.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
 CC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RI;
 RX MEDLINE-20036896; PubMed-10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.,
 RT "Genome sequence of the radioresistant bacterium *Deinococcus*
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL, AE001882; AAF09788.1; -;
 DR TIGR; DR0203; -;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 275 AA; 29820 MW; F17E96929ED9EAA1 CRC64;

Query Match 59.7%; Score 37; DB 16; Length 275;
 Best Local Similarity 85.7%; Pred. No. 58;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 8 AWACGXA 14
 |||||
 Db 163 AWACGSA 169

RESULT 12
 O9XAP4 PRELIMINARY: PRT; 334 AA.
 ID O9XAP4;
 AC O9XAP4;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Putative hydrolase.
 GN SC04552 OR SC016A.31.
 OS Streptomyces coelicolor.
 CC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-A3(2) / M145;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Horsley T., Howarth S.,
 RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.,
 RT "Complete genome sequence of the model actinomycete *Streptomyces*
coelicolor A3(2).";
 RL Nature 417:141-147(2002).
 DR EMBL, AL078618; CAB44541.1; -;
 KW Hydrolase.
 SQ SEQUENCE 334 AA; 33828 MW; 9789F3A06B587F1 CRC64;

Query Match 59.7% Score 37; DB 16; Length 334;
 Best Local Similarity 57.1% Pred. No. 71;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 1 GAXXXETAWACGX 14
 DB 115 GASGTETAAACGSA 128

RESULT 13

O9RH01 PRELIMINARY; PRT; 448 AA.
 AC O9RH01:
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
 DE Putative bile salt transporter.
 OS Lactobacillus acidophilus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Lactobacillaceae; Lactobacillus.
 OX NCBI_TaxID=1579;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KS-13;
 RA Moser S.A., Elkins C.A., Savage D.C.;
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF091248; AAF15542.1; --
 SQ SEQUENCE 448 AA: 48705 MW: 9BFC72C5B660C8AD CRC64;

Query Match 59.7% Score 37; DB 2; Length 448;
 Best Local Similarity 54.5% Pred. No. 95;
 Matches 6; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 1 GAXXXETAWAC 11
 DB 401 GATSTTTAMCC 411

RESULT 14

O8X4B7 PRELIMINARY; PRT; 502 AA.
 AC O8X4B7:
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Hypothetical protein z2116.
 GN z2116.
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Fosfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RL Nature 409:528-533(2001).
 DR EMBL: AE005348; AAG56186.1; --
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 502 AA: 56615 MW: 26451039C4EAB643 CRC64;

Query Match 59.7% Score 37; DB 16; Length 502;
 Best Local Similarity 42.9% Pred. No. 11e+02;
 Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 GAXXXETAWACGX 14

DB 178 GRDITDTKWRCCGA 191

RESULT 15

O8VNN7 PRELIMINARY; PRT; 670 AA.
 ID O8VNN7:
 AC O8VNN7:
 DT 01-MAR-2002 (TREMblrel. 20, Created)
 DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
 DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
 DE Putative terminase, large subunit.
 GN l32.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1639/77;
 RA Koehler B., Karch H., Schreier P., Schmidt H.;
 RT "Chromosomal integration sites of Shiga toxin-converting
 RT bacteriophages."
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ304858; CAC83155.1; --
 SQ SEQUENCE 670 AA: 76298 MW: D922035847CDA8E7 CRC64;

Query Match 59.7% Score 37; DB 2; Length 670;
 Best Local Similarity 42.9% Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

OY 1 GAXXXETAWACGX 14
 DB 178 GRDITDTKWRCCGA 191

Search completed: March 25, 2003, 08:21:35
 Job time : 32.9697 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:17:33 ; Search time 29.0909 Seconds
(without alignments)
45.805 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPLAAVKM 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_101002.*
1: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| 1 | 48 | 100.0 | 10 | 20 | AAV42753 |
| 2 | 48 | 100.0 | 10 | 20 | AAV39331 |
| 3 | 35 | 72.9 | 87 | 23 | ABP08699 |
| 4 | 35 | 72.9 | 132 | 22 | ABP07944 |
| 5 | 33 | 68.8 | 36 | 22 | AAW85180 |
| 6 | 32 | 66.7 | 56 | 23 | ABP02748 |
| 7 | 32 | 66.7 | 186 | 23 | ABP27852 |
| 8 | 32 | 66.7 | 323 | 23 | ABP91767 |
| 9 | 32 | 66.7 | 402 | 23 | ABP30120 |
| 10 | 31 | 64.6 | 59 | 23 | ABP26059 |

| | | | | | |
|----|----|------|-----|----|----------|
| 11 | 31 | 64.6 | 181 | 22 | AAW68929 |
| 12 | 31 | 64.6 | 185 | 22 | AAU30414 |
| 13 | 31 | 64.6 | 207 | 22 | AAW50648 |
| 14 | 31 | 64.6 | 315 | 23 | ABP60886 |
| 15 | 31 | 64.6 | 363 | 18 | AAW20581 |
| 16 | 31 | 64.6 | 363 | 18 | AAW24696 |
| 17 | 31 | 64.6 | 436 | 18 | AAW55473 |
| 18 | 31 | 64.6 | 436 | 18 | AAW55543 |
| 19 | 31 | 64.6 | 442 | 18 | AAW55250 |
| 20 | 31 | 64.6 | 509 | 20 | AAV44020 |
| 21 | 31 | 64.6 | 509 | 20 | AAW93575 |
| 22 | 31 | 64.6 | 509 | 23 | AAE21525 |
| 23 | 31 | 64.6 | 509 | 23 | ABB57092 |
| 24 | 31 | 64.6 | 707 | 21 | AAW32499 |
| 25 | 31 | 64.6 | 716 | 22 | ABB65655 |
| 26 | 31 | 64.6 | 823 | 19 | AAW98507 |
| 27 | 30 | 62.5 | 51 | 20 | AAV36026 |
| 28 | 30 | 62.5 | 52 | 21 | AAW34318 |
| 29 | 30 | 62.5 | 72 | 23 | ABP02946 |
| 30 | 30 | 62.5 | 76 | 22 | AAU42162 |
| 31 | 30 | 62.5 | 113 | 23 | ABB50028 |
| 32 | 30 | 62.5 | 168 | 22 | AAW89204 |
| 33 | 30 | 62.5 | 195 | 22 | AAO13539 |
| 34 | 30 | 62.5 | 236 | 22 | AAW81256 |
| 35 | 30 | 62.5 | 239 | 21 | AAW01418 |
| 36 | 30 | 62.5 | 258 | 22 | ABB70394 |
| 37 | 30 | 62.5 | 266 | 21 | AAW14127 |
| 38 | 30 | 62.5 | 308 | 21 | AAW88300 |
| 39 | 30 | 62.5 | 315 | 23 | ABB55045 |
| 40 | 30 | 62.5 | 319 | 21 | AAW88296 |
| 41 | 30 | 62.5 | 335 | 22 | AAW91138 |
| 42 | 30 | 62.5 | 343 | 21 | AAW88281 |
| 43 | 30 | 62.5 | 343 | 22 | AAW81263 |
| 44 | 30 | 62.5 | 343 | 22 | AAW88425 |
| 45 | 30 | 62.5 | 354 | 21 | AAW88297 |

ALIGNMENTS

| | |
|--|--|
| RESULT 1 | |
| AAV42753 | |
| AAV42753 standard; peptide: 10 AA. | |
| AAV42753: | |
| 20-DEC-1999 (first entry) | |
| Wheat amyloplast ADP-glucose transporter peptide #1. | |
| Starch biosynthesis: amyloplast; ADP-glucose: transport; import; | |
| amylopectin; amylose; branching; chemical structure; transgenic plant; | |
| optimisation; Industrial applications. | |
| XX Trifolium aestivum. | |
| OS WO9947682-A1. | |
| PN MO9947682-A1. | |
| PD 23-SEP-1999. | |
| XX 19-MAR-1999; 99WO-GB00728. | |
| PF 20-MAR-1998; 98GB-0005939. | |
| PR (UYMA-) UNIV VICTORIA MANCHESTER. | |
| XX Emes MU, Tellow IJ, Bowsher CG; | |
| DR WPI: 1999-590977/50. | |
| XX New transporter complex protein useful for modulating starch content in | |
| PT plants, especially useful in food production - | |
| XX | |

PS Claim 1; Page 3; 28pp; English.

XX This sequence represents a wheat amyloplast ADP-glucose transporter

CC peptide, #1. The wheat amyloplast ADP-glucose transporter is associated

CC with the amyloplast membrane and comprises at least two proteins; this

CC sequence, along with peptides #2 (AA142754) and #3 (AA142755) are

CC components of one of these proteins. Peptides #4-#7 (AA142756-142759)

CC represent sequences within the second protein (AA142760). The sugar

CC nucleotide ADP-glucose is the immediate substrate for starch synthesis,

CC which occurs in the amyloplast; however, ADP-glucose is mainly

CC synthesised outside the amyloplast in the cytoplasm. The ADP-glucose

CC transporter is responsible for the import of ADP-glucose into the

CC amyloplast and therefore plays a pivotal role in the regulation of starch

CC synthesis. The transporter not only influences starch yield, but also

CC quality as the starch synthases involved in amylose and amylopectin

CC synthesis have different affinities for ADP-glucose. Variations in the

CC chemical structure of starch are determined by the ratio of amylose to

CC amylopectin, and by the degree of branching in amylopectin in the starch

CC polymer. These variations can significantly alter the properties of

CC starch. The ADP-glucose transporter complex is useful for generating

CC transgenic plants in which the starch quality can be optimised for

CC industrial applications in which starch is used. For example, transgenic

CC plants which have an increased amylose content in starch are useful for

CC production of starch with increased viscosity and gel strength, which

CC prevents baked food going stale so quickly. Conversely, an increased

CC amylopectin content in the starch produces waxy starch useful as

CC thickening agents in food and coatings.

XX

SO Sequence 10 AA;

Query Match 100.0%; Score 48; DB 20; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00038;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SMPLNAAVKM 10

Db 1 SMPLNAAVKM 10

RESULT 2

AA139331

ID AAY39331 standard; peptide: 10 AA.

XX

AC AAY39331;

XX

DT 01-DEC-1999 (first entry)

XX

DE ADP glucose transporter peptide 1.

XX

KW ADP glucose transporter; transform plant cell; wheat; starch production;

KW waxy starch; thickening agent; food; coating; increased viscosity; stale;

KW gel strength; baked food.

XX

OS Triticum aestivum.

XX

PN WO9947681-A1.

XX

PD 23-SEP-1999.

XX

PE 19-MAR-1999; 99WO-GB00727.

XX

PR 20-MAR-1998; 98GB-0005939.

XX

PA (UYMA-) UNIV VICTORIA MANCHESTER.

XX

PI Emes MJ, Tetlow IJ, Bowers CG;

XX

DR WPI: 1999-571841/48.

XX

PT ADP glucose transporter protein used for modifying plant starch

PT production

XX

PS Claim 1; Page 15; 26pp; English.

XX Peptides AAY39331-Y39337 are fragments of an ADP glucose transporter

CC protein. The protein contains at least one of the peptide sequences, and

CC is capable of ADP glucose transport. A DNA molecule encoding an ADP

CC glucose transporter protein can be used to transform plant cells. The

CC ADP glucose transporter protein can be used to regulate starch

CC production from a plant. The plants can then be used to produce waxy

CC starches that can be used as thickening agents in food and coatings.

CC Alternatively the plants can be used to produce a starch with increased

CC viscosity and gel strength, which can be incorporated in baked food which

CC stays fresh for a longer length of time.

XX

SO Sequence 10 AA;

Query Match 100.0%; Score 48; DB 20; Length 10;

Best Local Similarity 100.0%; Pred. No. 0.00038;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 SMPLNAAVKM 10

Db 1 SMPLNAAVKM 10

RESULT 3

ABP08699

ID ABP08699 standard; Protein; 87 AA.

XX

AC ABP08699;

XX

DT 24-JUN-2002 (first entry)

XX

DE Human ORFX protein sequence SEQ ID NO:17380.

XX

KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;

KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;

KW degenerative disorder; osteoarthritis; neurodegenerative disorder;

KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;

KW hypertension; hypothyroidism; cholesterol ester storage disease;

KW immune deficiency; immune disorder; infectious disease;

KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;

KW myasthenia gravis.

XX

OS Homo sapiens.

XX

PN WO200192523-A2.

XX

PD 06-DEC-2001.

XX

PE 29-MAY-2001; 2001WO-US10836.

XX

PR 30-MAY-2000; 2000US-206132P.

XX

PR 29-AUG-2000; 2000US-228716P.

XX

PA (CURA-) CURAGEN CORP.

XX

PI Shimkets RA, Leach MD;

XX

DR WPI: 2002-106308/14.

XX

DR N-PSDB; ABN24451.

XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,

PT preventing and treating cardiovascular disease, neurodegenerative,

PT hyperproliferative disorders and autoimmune disorders

XX

PS Disclosure; SEQ ID 17380; 1037pp; English.

XX

CC The present invention describes substantially purified human proteins

CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1

CC in the specification)). ABN15762 to ABN27252 encode the human ORFX

CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for

CC treating or preventing a pathology associated with an ORFX-associated

CC disorder in humans, and in the manufacture of a medicament for treating a

CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

| | | | |
|----|----|----------|----|
| QY | 1 | SMPLNAAY | 8 |
| | : | | |
| Db | 37 | TMPLNAAY | 44 |

| | |
|----|---------------------------|
| AC | ABP07944; |
| XX | |
| DT | 25-JUN-2002 (first entry) |

Human ORFX protein sequence SEQ ID NO:15870.

KM Human: open reading frame: ORFX gene therapy; cancer; cirrhosis;
KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KM degenerative disorder: osteoarthritis; neurodegenerative disorder;
KM cardiovascular disease: diabetes mellitus; systemic lupus erythematosus;
KM hypertension; hypothyroidism; cholesterol ester storage disease;
KM immune deficiency; immune disorder; infectious disease;
KM autoimmune disorder: rheumatoid arthritis; autoimmune thyroiditis;
KM myasthenia gravis.

OS Homo sapiens.

PN WO200192523-A2

PD 06-DEC-2001.

PF 29-MAY-2001; 2001WO-US10836

PR 30-MAY-2000; 2000US-206132P

XX

XX
XX

XX

DR N-PSDB; ABN23696.

PT Novel human polypeptides and polynucleotides useful for diagnosing,

hyperproliferative disorders and autoimmune disorders

PS Disclosure; SEQ ID 15870; 1037pp; English

XX

The present invention describes substantially purified human proteins (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1 in the specification)). ABN15762 to ABN27552 encode the human ORFX proteins given in ABP00010 to ABP11500. ORFX proteins are useful for treating or preventing a pathology associated with an ORFX-associated disorder in humans, and in the manufacture of a medicament for treating a syndrome associated with ORFX-associated disorder. ORFX polynucleotide sequences can be used in gene therapy. ORFX sequences can be used in the treatment of cancer, hyperproliferative disorders, cirrhosis of liver, porosis, benign tumours, keloid, degenerative disorders, haemorrhage, osteoarthritis, neurodegenerative disorders, disorders related to organ transplantation, cardiovascular diseases, diabetes mellitus, systemic lupus erythematosus, hypertension, hypothyroidism, cholesterol ester storage disease, various immune deficiencies and disorders, infectious diseases, autoimmune disorders such as multiple sclerosis, rheumatoid arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host disease and autoimmune inflammatory eye disease. ORFX proteins are also useful for treating burns, incisions, ulcers, for treating osteoporosis, bone degenerative disorders, or periodontal disease, and for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues and conditions resulting from systemic cytokine damage.

N.B. The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pat_sequences](http://wipo.int/pub/published_pat_sequences).

Sequence 132 AA;

| | | | | |
|-------------|------------------|----------------|--------------|------------|
| Query Match | 72.9% | Score 35 | DB 23 | Length 132 |
| Best Local | Similarity 60.0% | Pred No. 6.7 | | |
| Matches | 6 | Conservative 3 | Mismatches 1 | Indels 0 |
| | | | Gaps | 0 |

| | | | |
|----|---|-----------|----|
| Qy | 1 | SMPLNAVKM | 10 |
| | | : : : | |
| Db | 2 | SIPMNCVKM | 11 |

| | |
|----------|-----------------------------------|
| RESULT 5 | |
| AAM85180 | |
| ID | AAM85180 standard; Protein; 36 AA |

| | |
|----|---------------------------|
| AC | AAM85180; |
| XX | |
| DT | 07-NOV-2001 (first entry) |

| DE | Human immune/haematopoietic antigen SEQ | ID NO:12773 |
|-----|---|-------------|
| ... | | |

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer,

XX XX

XX
XX

XX
00-AUG-2007

| | | | | |
|----|----|------|--------|---------|
| XX | 17 | 2001 | 200110 | 2001254 |
| XX | 17 | 2001 | 200110 | 2001254 |

XX
3

PR 04-FEB-2000; 2000US-0180628

PR 02-MAR-2000; 2000US-0186350

PR 17-MAR-2000; 2000US-0190076

PR 19-MAY-2000; 2000US-0205515

PR 28-JUN-2000; 2000US-0214886

PR 07-JUL-2000; 2000US-0216647

PR 11-JUL-2000; 2000US-0217487

PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233053.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PR XX (HUMA-) HUMAN GENOME SCI INC.
PR PA
PR PA
PR XX
PR PI
PR XX
PR DR WPI; 2001-483426/52.
PR DR N-PSDB; AAK57961.
PR PT
PR PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PR PT useful for preventing, diagnosing and/or treating cancers and
PR PT metastasis -
PR XX
PR PS
PR PS
PR XX
PR CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
PR CC amino acid sequences given in AAMW82170 to AAMW1921. (I) have cytostatic
PR CC activity and can be used in gene therapy and vaccine production. (I)
PR CC proteins and polynucleotides may be used in the prevention, diagnosis and
PR CC treatment of diseases associated with inappropriate (I) expression. For
PR CC example, they may be used to treat disorders associated with decreased
PR CC expression by rectifying mutations or deletions in a patient's genome
PR CC that affect the activity of (I) by expressing inactive proteins or to
PR CC supplement the patients own production of (I). Additionally, (I)
PR CC polynucleotides may be used to produce the secreted (I), by inserting
PR CC the nucleic acids into a host cell and culturing the cell to express the

CC protein. (1) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK7654 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54542 to AAK54950 and AAK62169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 36 AA;
Query Match 68.8%; Score 33; DB 22; Length 36;
Best Local Similarity 66.7%; Pred. No. 4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 2 MPLNAVKM 10
||| |
DB 2 MPLNCTIKM 10
RESULT 6
ABP02748
ID ABP02748 standard; Protein: 56 AA.
XX
AC ABP02748;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5478.
XX
KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KM hyperproliferative disorder; psoriasis; benign tumor; hemorrhage;
KM degenerative disorder; osteoarthritis; neurodegenerative disorder;
KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KM hypertension; hypothyroidism; cholesterol ester storage disease;
KM immune deficiency; immune disorder; infectious disease;
KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KM myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
XX
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shmkets RA, Leach MD;
XX
XX WPI: 2002-106308/14.
DR N-PSDB; ABN18500.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders -
XX
PS Disclosure: SEQ ID 5478; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC syndrome in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used in gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumors, keloid, degenerative disorders, hemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic

CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 56 AA;
Query Match 66.7%; Score 32; DB 23; Length 56;
Best Local Similarity 60.0%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 SMPLNAVKM 10
:||| |:
DB 24 NMPLVNMPM 33
RESULT 7
ABP27852
ID ABP27852 standard; Protein: 186 AA.
XX
AC ABP27852;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 4880.
XX
KM Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
KM group A streptococcus; Streptococcus pyogenes; antibacterial;
KM antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN WO200234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
XX
PR 24-NOV-2000; 2000GB-0028727.
XX
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Maignani V, Margarit Ros VI, Grandi G, Fraser C;
PI Tettelin H;
XX
XX WPI: 2002-352536/38.
DR N-PSDB; ABN68483.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 1; Page 3650; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
CC activity. (1), nucleic acids encoding (1), ABN6044-ABN71526 and
CC antibodies that bind (1) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 SO Sequence 186 AA;

Query Match 66.7%; Score 32; DB 23; Length 186;
 Best Local Similarity 60.0%; Pred. No. 48;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMPLNAAVKM 10
 I:|I |I: |
 Db 70 SIPLGAATNM 79

RESULT 8

ABB91767
 ID ABB91767 standard; Protein: 323 AA.

XX ABB91767;

DT 31-MAY-2002 (first entry)

XX Herbicidally active polypeptide SEQ ID NO 978.

XX Herbicidal; plant; agriculture; herbicide.

XX Arabidopsis thaliana.

XX WO200210210-A2.

XX 07-FEB-2002.

XX 28-AUG-2001; 2001WO-EP09892.

XX 28-AUG-2001; 2001WO-EP09892.

XX (FARB) BAYER AG.

XX Tietjen K, Weidner M;

XX WPI: 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -

PS Claim 5; SEQ ID NO 978; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

XX Sequence 323 AA;

Query Match 66.7%; Score 32; DB 23; Length 323;
 Best Local Similarity 75.0%; Pred. No. 92;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVK 9
 |I|I |I:
 Db 74 MPLNAVKR 81

RESULT 9

ABP30120
 ID ABP30120 standard; Protein: 402 AA.

XX ABP30120;

DT 02-JUL-2002 (first entry)

XX Streptococcus polypeptide SEQ ID NO 9416.

XX Streptococcus GAS; group B streptococcus; Streptococcus agalactiae;
 XX group A streptococcus; Streptococcus pyogenes; antibacterial;
 XX anti-inflammatory; infection; vaccine; meningitis; gene therapy.

XX Streptococcus agalactiae.

XX WO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelin H;

XX WPI: 2002-352536/38.

XX N-PSDB; ABN70751.

XX New Streptococcus protein for the treatment or prevention of infection
 XX or disease caused by Streptococcus bacteria, such as meningitis, and
 XX for detecting a compound that binds to the protein -

PS Claim 1; Page 4067-4068; 4525pp; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.

XX Sequence 402 AA;

Query Match 66.7%; Score 32; DB 23; Length 402;
 Best Local Similarity 60.0%; Pred. No. 1.2e+02;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMPLNAAVKM 10
 I:|I |I: |
 Db 286 SIPLGAATNM 295

RESULT 10

ABP26059 standard; Protein: 59 AA.

ABP26059;

02-JUL-2002 (first entry)

Streptococcus polypeptide SEQ ID NO 1294.

Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

group A streptococcus; Streptococcus pyogenes; antibacterial;

antimicrobial; infection; vaccine; meningitis; gene therapy.

Streptococcus agalactiae.

WO200234771-A2.

02-MAY-2002.

29-OCT-2001; 2001WO-GB04789.

27-OCT-2000; 2000GB-0026333.

24-NOV-2000; 2000GB-0028727.

07-MAR-2001; 2001GB-0005640.

(CHIR-) CHIRON SPA.

(GENO-) INST GENOMIC RES.

Telford J, Maignan V, Margalit Ros YI, Grandi G, Fraser C;

Tettelin H;

WPI: 2002-352536/38.

N-PSDB: ABN6690.

New Streptococcus protein for the treatment or prevention of infection

or disease caused by Streptococcus bacteria, such as meningitis, and

for detecting a compound that binds to the protein -

Claim 1; Page 3286; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B

Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

(Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in

the specification. The proteins have antibacterial and anti-inflammatory

activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and

antibodies that bind (I) are used in the manufacture of medicaments for

the treatment or prevention of infection or disease caused by

Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

Nucleic acids encoding (I) are used to detect Streptococcus in a

biological sample. (I) is used to determine whether a compound binds to

(I). A composition comprising (I) or a nucleic acid encoding (I), may be

used as a vaccine or diagnostic composition. The disease caused by

Streptococcus that is prevented or treated may be meningitis. Nucleic

acid encoding (I) may be used to recombinantly produce (I) and may be

used in gene therapy. Antibodies to (I) are used for affinity

chromatography, immunoassays, and distinguishing/identifying

Streptococcus proteins.

Sequence 59 AA:

Query Match 64.6%; Score 31; DB 23; Length 59;

Best Local Similarity 71.4%; Pred. No. 20;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

3 PLNAVK 9

1:|||||

5 PLNAVK 11

RESULT 11

AAB68929

AAB68929 standard; Protein: 181 AA.

AAB68929;

18-APR-2001 (first entry)

Neisseria meningitidis protein #28.

Meningococcus; meningitis; bacteraemia; vaccine; dba; fhaB; fhaA;

rn15; rfh; to1C.

Neisseria meningitidis.

EPI069133-A1.

17-JAN-2001.

13-JUL-1999; 99EP-0401764.

13-JUL-1999; 99EP-0401764.

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

(PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

Nassif X, Tinsley C;

WPI: 2001-082916/10.

N-PSDB: AAF56469.

Immunogenic polypeptides derived from Neisseria meningitidis and the

nucleic acids that encode them, useful for diagnosing and vaccinating

against Neisseria infections e.g. bacteraemia and meningitis -

Claim 3; Fig 28B; 240pp; English.

The present invention provides the protein and coding sequences of

several genes from Neisseria meningitidis. These include the dba, fhaB,

fhaA, rnf5, rnf17, rnf18, rnf19, rnf20, rnf21 and to1C genes. These can

be used in the diagnosis and treatment of infection by the bacterium,

which can lead to meningitis and bacteraemia, and in vaccines to prevent

such infection.

Sequence 181 AA:

Query Match 64.6%; Score 31; DB 22; Length 181;

Best Local Similarity 60.0%; Pred. No. 77;

Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

1 SMPLNAAVK 10

|||||

57 NMPLREAVKL 66

RESULT 12

AAU30414 standard; Protein: 185 AA.

AAU30414;

18-DEC-2001 (first entry)

Novel human secreted protein #905.

Human: vaccination; gene therapy; nutritional supplement;

stem cell proliferation; haematopoiesis; nerve tissue regeneration;

immune suppression; immune stimulation; anti-inflammatory; leukaemia.

Homo sapiens.

WO200179449-A2.

25-OCT-2001.

XX 16-APR-2001; 2001WO-US08656.
 XX 18-APR-2000; 2000US-0552929.
 PR 26-JAN-2001; 2001US-0770160.
 XX (HXSE-) HXSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI: 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20; Page 291; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU3304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SO Sequence 185 AA:
 OY 2 MPLNAVK 10
 11111111
 Db 42 MPLNGYLM 50
 Query Match 64.6%; Score 31; DB 22; Length 185;
 Best Local Similarity 66.7%; Pred. No. 79;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 RESULT 13
 AAM50648
 ID AAM50648 standard; Protein; 207 AA.
 AC AAM50648;
 XX
 DT 04-APR-2002 (first entry)
 XX
 DE Arabidopsis BAL:BAP-like protein BAL.
 XX
 KW BAL:BAP-like protein; growth; homeostasis; thermotolerance;
 KW transgenic plant; plant.
 XX
 OS Arabidopsis thaliana.
 OS
 PN WO200200697-A2.
 XX
 PD 03-JAN-2002.
 XX
 PF 25-JUN-2001; 2001WO-US20172.
 XX
 PR 23-JUN-2000; 2000US-213863P.
 XX
 PA (WHED) WHITEHEAD INST BIOMEDICAL RES.
 XX
 PI Hua J, Grisafi P, Fink GR;
 XX

DR WPI: 2002-139899/18.
 DR N-PSDB: ABA91260.
 XX
 PT New phospholipid binding proteins and nucleic acids, useful for
 PT modulating plant growth homeostasis, controlling cell expansion and
 PT cell division, or producing plants where larger fruits and increased
 PT biomass are desired -
 XX
 PS Disclosure: Fig 12B; 78pp; English.
 XX
 CC The present sequence is that of the predicted protein product of
 CC the BAL (BAL:BAP-like) protein gene of Arabidopsis thaliana. The
 CC BAL protein is necessary for normal plant growth. The invention is
 CC directed to isolated B0N1, B0N2, B0N3, BAP1 and BAL nucleic acids
 CC (see ABA91256-60), which encode proteins (see AAM50644-48) that are
 CC necessary for normal growth, controlling cell expansion and cell
 CC division, and thereby affecting the size and rate at which the
 CC plant grows when exposed to lower temperatures. Transgenic plants
 CC are provided that are smaller than the wild-type as a result of
 CC inhibition of B0N1, B0N2, B0N3, BAP1 and/or BAL, especially
 CC angiosperms and gymnosperms, ornamental plants and turfgrass.
 CC Transgenic plants are also provided that are larger than the
 CC wild-type as a result of enhancement of B0N1, B0N2, B0N3, BAP1
 CC and/or BAL, especially crop plants and biomass plants. Modulation
 CC of these genes provides increased yield, or growth at a higher
 CC altitude or lower temperature.
 CC
 SQ Sequence 207 AA:
 OY 2 MPLNAVK 9
 11111111
 Db 64 MPINMSVR 71
 Query Match 64.6%; Score 31; DB 23; Length 207;
 Best Local Similarity 62.5%; Pred. No. 91;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 RESULT 14
 ABP60886
 ID ABP60886 standard; Protein; 315 AA.
 AC ABP60886;
 XX
 DT 06-SEP-2002 (first entry)
 XX
 DE Bacillus halodurans thioedoxin reductase SEQ ID NO:235.
 XX
 KW Multimeric protein; redox protein; thioedoxin; thioedoxin reductase;
 KW oil body; ophthalmological; antidiabetic; cytosolic; antiporiatric;
 KW vasotropic; vulnary; antibacterial; immunosuppressive; antiulcer;
 KW food product; milk; wheat; oxidative stress; cataract; diabetes;
 KW chronic obstructive pulmonary disease; envenomation; psoriasis; sepsis;
 KW bronchiopulmonary disease; malignancy; reperfusion injury; wound healing;
 KW gastro intestinal bleeding; intestinal bowel disease; ulcer;
 KW gastro oesophageal reflux disease.
 XX
 OS Bacillus halodurans.
 OS
 PN WO200250289-A1.
 XX
 PD 27-JUN-2002.
 XX
 PF 19-DEC-2001; 2001WO-US50240.
 XX
 PR 19-DEC-2000; 2000US-0742900.
 XX
 PR 05-JUL-2001; 2001US-302885P.
 PR 04-DEC-2001; 2001US-006038.
 XX
 PA (SEMB-) SEMBIOSYS GENETICS INC.
 XX
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.
 XX
 PI Van Rooijen G, Deckers H, Helfeltz PB, Briggs SP, Dalmia BK;
 XX

PI Del Val G, Zaplachinski S, Moloney M;
 XX
 DR MPI: 2002-508806/54.
 XX
 PT Producing oil body associated with recombinant multimeric protein
 PT complex e.g. redox proteins and immunoglobulins comprises producing
 PT recombinant polypeptides capable of forming the complex in cells
 PT comprising oil bodies -
 XX
 PS Claim 82: Page 296-297; 362pp: English.
 XX
 CC The present invention describes a method (M1) for producing an oil body
 CC associated with a recombinant multimeric protein complex (MPC). M1
 CC comprises producing in a cell comprising oil bodies a first and second
 CC recombinant polypeptide (P1, P2), where P1 is capable of associating
 CC with P2 to form the MPC and associating the complex with an occlusion
 CC body (OB) through an OB-pore-forming protein capable of associating with OB
 CC and P1. M1 is useful for producing an oil body associated with a
 CC recombinant MPC. The oil bodies are further formulated for use in the
 CC preparation of a food product such as milk or wheat based food product,
 CC personal care product which reduces the oxidative stress on the surface
 CC area of the human body or used to lighten the skin, or a pharmaceutical
 CC composition used to treat chronic obstructive pulmonary disease (COPD),
 CC cataracts, diabetes, emphysema, bronchiopulmonary disease (BPD),
 CC malignancies, reperfusion injury, wound healing, sepsis, gastro
 CC intestinal (GI) bleeding, intestinal bowel disease (IBD), ulcers, GERD
 CC (gastro oesophageal reflux disease). ABN89569 to ABN89593 and ABP60677
 CC to APP60964 represent sequence given in the exemplification of the
 CC present invention.
 XX
 SQ Sequence 315 AA;
 XX
 Query Match 64.6%; Score 31; DB 23; Length 315;
 Best Local Similarity 75.0%; Pred. No. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 MPLNAVK 9 :|||||
 DB 243 LPLNEAVK 250
 RESULT 15
 ID AAM20581
 AC AAM20581 standard; protein; 363 AA.
 XX
 AC AAM20581:
 XX
 DT 04-JUN-1997 (first entry)
 XX
 DE Helicobacter pylori cell envelope protein, 978477.aa.
 XX
 KW Cell envelope; vaccine; prevention; treatment; infection;
 KW identification; binding compound; bacterium; life cycle;
 KW activator; bacteria; inhibitor; duodenal ulcer disease;
 KW chronic gastritis; diagnosis.
 XX
 OS Helicobacter pylori.
 XX
 FH key Location/Qualifiers
 FT Misc-difference 11 /note= "unknown"
 FT Misc-difference 337 /note= "unknown"
 FT Misc-difference 339 /note= "unknown"
 FT Misc-difference 339 /note= "unknown"
 FT Misc-difference 356 /note= "unknown"
 FT Misc-difference 360 /note= "unknown"
 FT Misc-difference 363 /note= "unknown"
 FT Misc-difference 363 /note= "unknown"
 XX
 PN WO9640893-A1.

XX
 PD 19-DEC-1996.
 XX
 PF 06-JUN-1996; 96WO-US09122.
 XX
 PR 01-APR-1996; 96US-0630405.
 XX
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 XX
 PI Berglindh OT, Smith D, Mellgaerd BL;
 XX
 DR MPI: 1997-052306/05.
 DR N-PSDB; AAT67834.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 56: Pages 735-736; 1481pp: English.
 XX
 CC The present sequence is a Helicobacter pylori cell envelope
 CC protein, which may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors.
 CC The genomic sequence of H. pylori (ATCC 55679) was determined from
 CC overlapping contigs generated by mechanically shearing the
 CC bacterial DNA. The sequences were analysed for ORF of at least 180
 CC nucleotides, and the predicted coding regions defined by computer
 CC evaluation. To identify likely H. pylori antigens for vaccine
 CC development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences
 CC of interest, particular regions can be isolated from H. pylori by
 CC PCR amplification for recombinant polypeptide production, e.g. in
 CC E. coli hosts.
 CC A 100 microg dose of the protein was given to female SPF BALB/c
 CC mice infected with a H. pylori type 2 strain. The mice were also
 CC given 10 microg of cholera toxin as adjuvant. Omeprazole
 CC (400 micromol/kg) was given orally 3-5 hours prior to immunisation
 CC to protect the protein from degradation. The mice were sacrificed
 CC 2-4 weeks after final immunisation and their serum antibody titre
 CC determined to be about 400, compared to about 350 for a LacZ
 CC control.
 CC
 XX
 SQ Sequence 363 AA;
 XX
 Query Match 64.6%; Score 31; DB 18; Length 363;
 Best Local Similarity 75.0%; Pred. No. 1.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SMPLNAAV 8 :|||||
 DB 257 SMPLNAAV 264
 Search completed: March 25, 2003, 08:19:40
 Job time : 31.0909 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:21:42 ; Search time 10.6061 Seconds
(without alignments)
27.742 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPLNAAVKM 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PT05.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|---------------------|
| 1 | 31 | 64.6 | 509 | 2 | US-08-559-505-4 |
| 2 | 31 | 64.6 | 509 | 2 | US-08-749-907-4 |
| 3 | 31 | 64.6 | 509 | 4 | US-09-241-581B-8 |
| 4 | 31 | 64.6 | 509 | 5 | PCR-US95-07721-8 |
| 5 | 30 | 62.5 | 291 | 2 | US-08-701-191A-22 |
| 6 | 30 | 62.5 | 410 | 2 | US-09-134-001C-4135 |
| 7 | 30 | 62.5 | 1135 | 4 | US-08-469-537A-97 |
| 8 | 30 | 62.5 | 1138 | 1 | US-08-323-474-8 |
| 9 | 30 | 62.5 | 1138 | 2 | US-08-469-537A-98 |
| 10 | 30 | 62.5 | 1138 | 2 | US-08-220-240A-5 |
| 11 | 29 | 60.4 | 210 | 4 | US-09-247-155-121 |
| 12 | 29 | 60.4 | 263 | 2 | US-08-892-690-3 |
| 13 | 29 | 60.4 | 264 | 3 | US-08-924-570A-2 |
| 14 | 29 | 60.4 | 528 | 3 | US-08-688-988-6 |
| 15 | 28 | 58.3 | 169 | 4 | US-09-311-626B-20 |
| 16 | 28 | 58.3 | 236 | 1 | US-08-266-570A-2 |
| 17 | 28 | 58.3 | 236 | 1 | US-08-266-570A-4 |
| 18 | 28 | 58.3 | 236 | 1 | US-08-266-570A-6 |
| 19 | 28 | 58.3 | 236 | 2 | US-08-757-046A-14 |
| 20 | 28 | 58.3 | 236 | 3 | US-09-447-208-14 |
| 21 | 28 | 58.3 | 236 | 4 | US-09-135-988-14 |
| 22 | 28 | 58.3 | 236 | 4 | US-09-277-716-14 |
| 23 | 28 | 58.3 | 236 | 4 | US-08-597-274A-14 |
| 24 | 28 | 58.3 | 236 | 4 | US-08-908-909-14 |
| 25 | 28 | 58.3 | 236 | 4 | US-09-609-161B-14 |
| 26 | 28 | 58.3 | 236 | 4 | US-08-990-103-14 |
| 27 | 28 | 58.3 | 249 | 2 | US-09-154-802-1 |

| | | | | | | |
|----|----|------|-----|---|------------------|------------------|
| 28 | 28 | 58.3 | 249 | 3 | US-09-373-029-1 | Sequence 1, Appl |
| 29 | 28 | 58.3 | 278 | 3 | US-08-491-954-2 | Sequence 2, Appl |
| 30 | 28 | 58.3 | 292 | 4 | US-09-224-048A-2 | Sequence 2, Appl |
| 31 | 28 | 58.3 | 346 | 2 | US-08-401-068-8 | Sequence 8, Appl |
| 32 | 28 | 58.3 | 346 | 2 | US-08-846-338-8 | Sequence 8, Appl |
| 33 | 28 | 58.3 | 346 | 3 | US-08-411-768B-2 | Sequence 2, Appl |
| 34 | 28 | 58.3 | 355 | 4 | US-09-390-131-5 | Sequence 5, Appl |
| 35 | 28 | 58.3 | 509 | 2 | US-08-559-505-2 | Sequence 2, Appl |
| 36 | 28 | 58.3 | 509 | 2 | US-08-749-907-2 | Sequence 2, Appl |
| 37 | 28 | 58.3 | 509 | 2 | US-08-890-980-2 | Sequence 2, Appl |
| 38 | 28 | 58.3 | 509 | 2 | US-08-890-980-4 | Sequence 2, Appl |
| 39 | 28 | 58.3 | 509 | 3 | US-08-890-979-2 | Sequence 2, Appl |
| 40 | 28 | 58.3 | 509 | 3 | US-08-890-979-4 | Sequence 2, Appl |
| 41 | 28 | 58.3 | 509 | 4 | US-09-032-894-2 | Sequence 2, Appl |
| 42 | 28 | 58.3 | 509 | 4 | US-09-032-894-4 | Sequence 2, Appl |
| 43 | 28 | 58.3 | 509 | 4 | US-09-031-626-2 | Sequence 2, Appl |
| 44 | 28 | 58.3 | 509 | 4 | US-09-031-626-4 | Sequence 4, Appl |
| 45 | 28 | 58.3 | 509 | 4 | US-09-241-581B-4 | Sequence 4, Appl |

ALIGNMENTS

RESULT 1
US-08-559-505-4
Sequence 4, Application US/08559505
Patent No. 5925333
GENERAL INFORMATION:
APPLICANT: Monty Krieger, Susan L. Acton, Attilio Rigotti, Helen H.
APPLICANT: Hobbs and Kathy Landshulz
TITLE OF INVENTION: METHODS FOR MODULATION OF LIPID UPTAKE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,505
FILING DATE: 29-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MIT7150
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ. ID NO. 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc. feature
LOCATION: 1..509
OTHER INFORMATION: /Function = "Amino acid sequence for the
murine Scavenger Receptor Class B1."
US-08-559-505-4
Query Match 64.6%; Score 31; DB 2; Length 509;
Best Local Similarity 55.6%; Pred. No. 16+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVKM 10
:1:1:111
Db 380 IPMNCVKM 388

RESULT 2

US-08-749-907-4
; Sequence 4, Application US/08749907
; Patent No. 5962322
; GENERAL INFORMATION:
; APPLICANT: Monty Krieger, Attilio Rigotti, and Karen Kozarsky
; TITLE OF INVENTION: METHODS FOR MODULATION OF CHOLESTEROL
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst, Arnall Golden & Gregory
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,907
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MIT7538
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 509 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: 1..509
; OTHER INFORMATION: /Function = "Amino acid sequence for the
; OTHER INFORMATION: murine Scavenger Receptor Class BI."
US-08-749-907-4

Query Match 64.6%; Score 31; DB 2; Length 509;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVKM 10
:1:1:111
Db 380 IPMNCVKM 388

RESULT 3

US-09-241-581B-8
; Sequence 8, Application US/09241581B
; Patent No. 6350859
; GENERAL INFORMATION:
; APPLICANT: Massachusetts Institute of Technology
; TITLE OF INVENTION: Class BI and CI Scavenger Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia

COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/241,581B
FILING DATE: 02-Feb-1999
CLASSIFICATION: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: MIT6620
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc. feature
LOCATION: 1..509
OTHER INFORMATION: /Function = "Amino acid sequence for
the murine Scavenger Receptor Class BI."
US-09-241-581B-8

Query Match 64.6%; Score 31; DB 4; Length 509;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVKM 10
:1:1:111
Db 380 IPMNCVKM 388

RESULT 4

PCT-US95-07721-8
; Sequence 8, Application PC/TUS9507721
; GENERAL INFORMATION:
; APPLICANT: Massachusetts Institute of Technology
; TITLE OF INVENTION: Class BI and CI Scavenger Receptors
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07721
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: MIT6620
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 509 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: misc. feature
LOCATION: 1..509
OTHER INFORMATION: /Function = "Amino acid sequence for
OTHER INFORMATION: the murine Scavenger Receptor Class B1."
PCT-US95-07721-8

Query Match 64.6%; Score 31; DB 5; Length 509;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
:|:|:|
DB 380 IPMNCVKM 388

RESULT 5
US-08-701-191A-22

Sequence 22, Application US/08701191A

Patent No. 5942428

GENERAL INFORMATION:

APPLICANT: Moosa Mohammadi, Joseph Schlessinger,

APPLICANT: and Steven R. Hubbard

TITLE OF INVENTION: CRYSTALS OF THE TYROSINE KINASE DOMAIN

TITLE OF INVENTION: OF NON-INSULIN RECEPTOR TYROSINE KINASE

NUMBER OF SEQUENCES: 41

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Lyon 6 Lyon

STREET: 633 West Fifth Street

STREET: Suite 4700

CITY: Los Angeles

STATE: California

COUNTRY: U.S.A.

ZIP: 90071-2066

COMPUTER READABLE FORM:

MEDIUM TYPE: 3.5" Diskette, 1.44 Mb

MEDIUM TYPE: storage

COMPUTER: IBM Compatible

OPERATING SYSTEM: IBM P.C. DOS 5.0

SOFTWARE: FASTSEQ for Windows 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/701.191A

FILING DATE: August 21, 1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warburg, Richard J.

REGISTRATION NUMBER: 32,327

REFERENCE/DOCKET NUMBER: 227/088

TELECOMMUNICATION INFORMATION:

TELEPHONE: (213) 489-1600

TELEFAX: (213) 955-0440

TELEX: 67-3510

INFORMATION FOR SEQ ID NO: 22:

SEQUENCE CHARACTERISTICS:

LENGTH: 291 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-701-191A-22

Query Match

Best Local Similarity

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
:|:|:|
DB 380 LKMNAAIKM 46

RESULT 6
US-09-134-001C-4135

Sequence 4135, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lyon Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/055,779

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 4135

LENGTH: 410

TYPE: PRT

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-4135

Query Match 62.5%; Score 30; DB 4; Length 410;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMPLNAAV 8
:|:|:|
DB 11 TMPLNAAV 18

RESULT 7
US-08-469-537A-97

Sequence 97, Application US/08469537A

Patent No. 5843749

GENERAL INFORMATION:

APPLICANT: Maisongier, et al.

TITLE OF INVENTION: EHR AND ROR TYROSINE

TITLE OF INVENTION: KINASES

NUMBER OF SEQUENCES: 107

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Regeneron Pharmaceuticals, Inc.

STREET: 777 Old Saw Mill River Road

CITY: Tarrytown

STATE: NY

COUNTRY: U.S.A.

ZIP: 10591

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/469,537A

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: USSN 08/406,247

FILING DATE: 17-MAR-1995

APPLICATION NUMBER: USSN 08/144,992

FILING DATE: 28-OCT-1993

APPLICATION NUMBER: USSN 07/736,559

FILING DATE: 26-JUL-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kempier, Ph.D., Gall M

REGISTRATION NUMBER: 32,143

REFERENCE/DOCKET NUMBER: REG 070C

TELECOMMUNICATION INFORMATION:

TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
INFORMATION FOR SEQ ID NO: 97:
SEQUENCE CHARACTERISTICS:
LENGTH: 1135 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-469-537A-97

Query Match 62.5%; Score 30; DB 2; Length 1135;
Best Local Similarity 55.6%; Pred. No. 4.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPLNAAYKM 10
: :|||:|
Db 860 LKMAAAYKM 868

RESULT 8
US-08-323-474-8
Sequence 8, Application US/08323474
Patent No. 5447860
GENERAL INFORMATION:
APPLICANT: Ziegler, Steven F.
TITLE OF INVENTION: NOVEL TYROSINE KINASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation
STREET: 51 University Street
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98101
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,474
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/905,600
FILING DATE: 26-JUN-1992
ATTORNEY/AGENT INFORMATION:
NAME: Seese, Kathryn A.
REGISTRATION NUMBER: 32,172
REFERENCE/DOCKET NUMBER: 2609
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 587-0430
TELEFAX: (206) 233-0644
TELEX: 756822
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1138 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-323-474-8

Query Match 62.5%; Score 30; DB 1; Length 1138;
Best Local Similarity 55.6%; Pred. No. 4.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPLNAAYKM 10
: :|||:|
Db 863 LKMAAAYKM 871

RESULT 9
US-08-469-537A-98
Sequence 98, Application US/08469537A
Patent No. 5843749
GENERAL INFORMATION:
APPLICANT: Maisongier, et al.
TITLE OF INVENTION: EHK AND FOR TYROSINE
TITLE OF INVENTION: KINASES
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: Regeneron Pharmaceuticals, Inc.
STREET: 777 Old Saw Mill River Road
CITY: Tarrytown
STATE: NY
COUNTRY: U.S.A.
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/469,537A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/406,247
FILING DATE: 17-MAR-1995
APPLICATION NUMBER: USSN 08/144,992
FILING DATE: 28-OCT-1993
APPLICATION NUMBER: USSN 07/736,559
FILING DATE: 26-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kempler, Ph.D., Gail M
REGISTRATION NUMBER: 32,143
REFERENCE/DOCKET NUMBER: REG 070C
TELECOMMUNICATION INFORMATION:
TELEPHONE: 914-345-7400
TELEFAX: 914-345-7721
TELEX:
INFORMATION FOR SEQ ID NO: 98:
SEQUENCE CHARACTERISTICS:
LENGTH: 1138 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-469-537A-98

Query Match 62.5%; Score 30; DB 2; Length 1138;
Best Local Similarity 55.6%; Pred. No. 4.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MPLNAAYKM 10
: :|||:|
Db 863 LKMAAAYKM 871

RESULT 10
US-08-220-240A-5
Sequence 5, Application US/08220240A
Patent No. 5955291
GENERAL INFORMATION:
APPLICANT: Alitalo, Kari
APPLICANT: Matikainen, Marja-Terttu
APPLICANT: Partanen, Juha
APPLICANT: Makiela, Tomi
APPLICANT: Korhonen, Jaana
TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE RECEPTOR
TITLE OF INVENTION: TYROSINE KINASE AND USES THEREOF
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:

```

ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606-6402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/220, 240A
FILING DATE: 29-MAR-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FI93/00006
FILING DATE: 08-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/817, 800
FILING DATE: 09-JAN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/167, 453
FILING DATE: 15-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Gass, David A.
REGISTRATION NUMBER: 38, 153
REFERENCE/DOCKET NUMBER: 29151/31958
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1138 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-220-240A-5

Query Match          62.5%; Score 30; DB 2; Length 1138;
Best Local Similarity 55.6%; Pred. No. 4.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY      2 MPLNAVKM 10
DB      863 LKMAAIKM 871

RESULT 11
US-09-247-155-121
Sequence 121, Application US/09247155A
Patent No. 6312922
GENERAL INFORMATION:
APPLICANT: Dumas Milne Edwards, Jean-Baptiste
APPLICANT: Ducleret, Aymeric
APPLICANT: Bouquellet, Lydie
TITLE OF INVENTION: Complementary DNAs
FILE REFERENCE: GENSET.021A
CURRENT APPLICATION NUMBER: US/09/247, 155A
CURRENT FILING DATE: 1999-02-09
EARLIER APPLICATION NUMBER: 60/074, 121
EARLIER FILING DATE: 1998-02-09
EARLIER APPLICATION NUMBER: 60/081, 563
EARLIER FILING DATE: 1998-04-13
EARLIER APPLICATION NUMBER: 60/096, 116
EARLIER FILING DATE: 1998-08-10
EARLIER APPLICATION NUMBER: 60/099, 273
EARLIER FILING DATE: 1998-10-04
NUMBER OF SEQ ID NOS: 182
SOFTWARE: Patent.pm
SEQ ID NO 121
LENGTH: 210
TYPE: PRT

```

```

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SIGNAL
LOCATION: -14...-1
US-09-247-155-121

Query Match          60.4%; Score 29; DB 4; Length 210;
Best Local Similarity 55.6%; Pred. No. 97;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 MPLNAVKM 10
DB      110 MPLNTSIWM 118

RESULT 12
US-08-892-690-3
Sequence 3, Application US/08892690
Patent No. 5932420
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
TITLE OF INVENTION: NEW INTEGRAL MEMBRANE PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/892, 690
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36, 749
REFERENCE/DOCKET NUMBER: PF-0339 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 624778
US-08-892-690-3

Query Match          60.4%; Score 29; DB 2; Length 263;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY      2 MPLNAVKM 10
DB      163 MPLNTSIWM 171

RESULT 13
US-08-924-570A-2

```

```

; Sequence 2, Application US/08924570A
; Patent No. 6093800
; GENERAL INFORMATION:
; APPLICANT: Reiter, Robert
; APPLICANT: Witter, Owen
; APPLICANT: Sawyer, Charles
; TITLE OF INVENTION: E25a PROTEIN, METHODS FOR
; TITLE OF INVENTION: PRODUCTION AND USE THEREOF
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Mandel & Adriano
; STREET: 725 Main Street
; CITY: Half Moon Bay
; STATE: CA
; COUNTRY: USA
; ZIP: 94019
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/924,570A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT US97/15606
; FILING DATE: 05-SEP-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Adriano, Sarah B
; REGISTRATION NUMBER: 34,470
; REFERENCE/DOCKET NUMBER: 30435.38W001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 213-258-5580
; TELEFAX: 213-254-1940
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 264 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
;
US-08-924-570A-2

Query Match
Best Local Similarity 60.4%; Score 29; DB 3; Length 264;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPLNAAVK 10
DB 164 MPLNTSIV 172

```

```

Query Match
Best Local Similarity 60.4%; Score 29; DB 3; Length 528;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SMPLNAAV 8
DB 388 SMPLTAAL 395

RESULT 15
US-09-311-626B-20
; Sequence 20, Application US/09311626B
; Patent No. 6399347
; GENERAL INFORMATION:
; APPLICANT: Jorgensen, Per Lina
; APPLICANT: Schmitt, Kirk
; APPLICANT: Andersen, Lene No. 6399347boe
; APPLICANT: Schuelein, Martin
; APPLICANT: Outtrup, Helle
; TITLE OF INVENTION: No. 6399347el Rhamnogalacturonan Hydrolases
; FILE REFERENCE: 5572.204-US
; CURRENT APPLICATION NUMBER: US/09/311,626B
; CURRENT FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: 0608/98
; PRIOR FILING DATE: 1998-05-01
; PRIOR APPLICATION NUMBER: 60/084,358
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 20
; LENGTH: 169
; TYPE: PRT
; ORGANISM: Bacillus agaradhaerens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 147
; OTHER INFORMATION: Xaa = any amino acid
;
US-09-311-626B-20

Query Match
Best Local Similarity 58.3%; Score 28; DB 4; Length 169;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 MPLNAAVK 9
DB 17 MPLNAAVR 24

Search completed: March 25, 2003, 08:22:58
Job time : 11.6061 secs

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:22:27 ; Search time 9.3934 Seconds
(Without alignments)
56.911 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPLNAAVKM 10

Scoring table:
BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 221153 seqs, 53462247 residues

Total number of hits satisfying chosen parameters: 221153

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----------------------|--------------------|
| 1 | 31 | 64.6 | 207 | 9 US-09-891-139A-11 | Sequence 11, Appl |
| 2 | 31 | 64.6 | 509 | 9 US-09-148-012-4 | Sequence 44, Appl |
| 3 | 31 | 64.6 | 823 | 9 US-09-895-913A-244 | Sequence 244, App |
| 4 | 30 | 62.5 | 168 | 10 US-09-731-872-324 | Sequence 324, App |
| 5 | 30 | 62.5 | 239 | 10 US-09-796-858-18 | Sequence 18, App |
| 6 | 30 | 62.5 | 335 | 9 US-09-738-626-4892 | Sequence 4892, App |
| 7 | 29 | 60.4 | 18 | 9 US-09-820-0568-5 | Sequence 5, Appl |
| 8 | 29 | 60.4 | 18 | 10 US-09-820-296-5 | Sequence 5, Appl |
| 9 | 29 | 60.4 | 263 | 9 US-10-174-590-184 | Sequence 184, App |
| 10 | 29 | 60.4 | 263 | 9 US-10-176-758-184 | Sequence 184, App |
| 11 | 29 | 60.4 | 263 | 9 US-10-175-727-184 | Sequence 184, App |
| 12 | 29 | 60.4 | 263 | 9 US-10-173-706-184 | Sequence 184, App |
| 13 | 29 | 60.4 | 263 | 9 US-10-175-738-184 | Sequence 184, App |
| 14 | 29 | 60.4 | 263 | 9 US-10-175-752-184 | Sequence 184, App |
| 15 | 29 | 60.4 | 263 | 9 US-10-176-482-184 | Sequence 184, App |
| 16 | 29 | 60.4 | 263 | 9 US-10-176-482-184 | Sequence 184, App |
| 17 | 29 | 60.4 | 263 | 9 US-10-176-757-184 | Sequence 184, App |
| 18 | 29 | 60.4 | 263 | 9 US-10-176-913-184 | Sequence 184, App |
| 19 | 29 | 60.4 | 263 | 9 US-10-180-552-184 | Sequence 184, App |
| | | | | 9 US-10-180-557-184 | Sequence 184, App |

| | | | | | |
|----|----|------|-----|---------------------|-------------------|
| 20 | 29 | 60.4 | 263 | 9 US-10-173-700-184 | Sequence 184, App |
| 21 | 29 | 60.4 | 263 | 9 US-10-174-572-184 | Sequence 184, App |
| 22 | 29 | 60.4 | 263 | 9 US-10-174-579-184 | Sequence 184, App |
| 23 | 29 | 60.4 | 263 | 9 US-10-174-582-184 | Sequence 184, App |
| 24 | 29 | 60.4 | 263 | 9 US-10-174-588-184 | Sequence 184, App |
| 25 | 29 | 60.4 | 263 | 9 US-10-175-739-184 | Sequence 184, App |
| 26 | 29 | 60.4 | 263 | 9 US-10-175-740-184 | Sequence 184, App |
| 27 | 29 | 60.4 | 263 | 9 US-10-175-743-184 | Sequence 184, App |
| 28 | 29 | 60.4 | 263 | 9 US-10-176-488-184 | Sequence 184, App |
| 29 | 29 | 60.4 | 263 | 9 US-10-176-492-184 | Sequence 184, App |
| 30 | 29 | 60.4 | 263 | 9 US-10-176-747-184 | Sequence 184, App |
| 31 | 29 | 60.4 | 263 | 9 US-10-176-750-184 | Sequence 184, App |
| 32 | 29 | 60.4 | 263 | 9 US-10-176-985-184 | Sequence 184, App |
| 33 | 29 | 60.4 | 263 | 9 US-10-176-987-184 | Sequence 184, App |
| 34 | 29 | 60.4 | 263 | 9 US-10-176-991-184 | Sequence 184, App |
| 35 | 29 | 60.4 | 263 | 9 US-10-176-992-184 | Sequence 184, App |
| 36 | 29 | 60.4 | 263 | 9 US-10-176-993-184 | Sequence 184, App |
| 37 | 29 | 60.4 | 263 | 9 US-10-184-658-184 | Sequence 184, App |
| 38 | 29 | 60.4 | 263 | 9 US-10-173-695-184 | Sequence 184, App |
| 39 | 29 | 60.4 | 263 | 9 US-10-173-697-184 | Sequence 184, App |
| 40 | 29 | 60.4 | 263 | 9 US-10-173-705-184 | Sequence 184, App |
| 41 | 29 | 60.4 | 263 | 9 US-10-174-576-184 | Sequence 184, App |
| 42 | 29 | 60.4 | 263 | 9 US-10-174-585-184 | Sequence 184, App |
| 43 | 29 | 60.4 | 263 | 9 US-10-174-586-184 | Sequence 184, App |
| 44 | 29 | 60.4 | 263 | 9 US-10-175-747-184 | Sequence 184, App |
| 45 | 29 | 60.4 | 263 | 9 US-10-176-481-184 | Sequence 184, App |

ALIGNMENTS

RESULT 1

US-09-891-139A-11
Sequence 11, Application US/09891139A
Publication No. US20020194639A1

GENERAL INFORMATION:

APPLICANT: Hua, Jian
APPLICANT: Grisaffi, Paula

TITLE OF INVENTION: Bonsai, A Phospholipid Binding Protein,
TITLE OF INVENTION: Is Required for Thermal Tolerance in Arabidopsis

FILE REFERENCE: 0399.2009-001
CURRENT APPLICATION NUMBER: US/09/891.139A

PRIOR FILING DATE: 2002-06-24
PRIOR APPLICATION NUMBER: US 60/213,863

NUMBER OF SEQ ID NOS: 12
SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 11
LENGTH: 207
TYPE: PRT

ORGANISM: Arabidopsis Thaliana

FEATURE:
NAME/KEY: PEPTIDE

LOCATION: (1)...(207)
OTHER INFORMATION: BAL1 Protein

US-09-891-139A-11

Query Match 64.6% Score 31: DB 9: Length 207;
Best Local Similarity 62.5% Pred. No. 51;
Matches 5; Conservative 3; Mismatches 0; Gaps 0;

QY 2 MPLNAAVK 9
DB 64 MPLNAAVK 71

US-09-148-012-4
Sequence 4, Application US/09148012

Patent No. US20020099040A1
GENERAL INFORMATION:
APPLICANT: Krieger, Monty

```

; TITLE OF INVENTION: SR-BI Antagonist And Use Thereof As Contraceptives And
; FILE REFERENCE: In The Treatment Of Steroidal Overproduction
; FILE REFERENCE: MIT7150CIP2
; CURRENT APPLICATION NUMBER: US/09/148,012
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: 60/057,943
; EARLIER FILING DATE: 1997-09-05
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 509
; TYPE: PRT
; ORGANISM: Mouse
; US-09-148-012-4

Query Match      64.6%; Score 31; DB 10; Length 509;
Best Local Similarity 55.6%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      2 MPLNAAYKM 10
      :| | | | |
Db      380 IPMNCVYKM 388

RESULT 3
US-09-895-913A-244
; Sequence 244, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 244
; LENGTH: 823
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-895-913A-244

Query Match      64.6%; Score 31; DB 9; Length 823;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      1 SMPLAAVK 9
      :| | | | |
Db      639 SPLPLSLIK 647

RESULT 4
US-09-731-872-324
; Sequence 324, Application US/09731872
; Patent No. US20020102604A1
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean Baptiste
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Jobert, Severin
; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS
; FILE REFERENCE: 78.US3.REG
; CURRENT APPLICATION NUMBER: US/09/731,872
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,629
; PRIOR FILING DATE: 1999-12-08
```

```

; PRIOR APPLICATION NUMBER: US 60/187,470
; PRIOR FILING DATE: 2000-03-06
; NUMBER OF SEQ ID NOS: 482
; SOFTWARE: Patent.pm
; SEQ ID NO 324
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -25...-1
; US-09-731-872-324

Query Match      62.5%; Score 30; DB 10; Length 168;
Best Local Similarity 55.6%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY      2 MPLNAAYKM 10
      :| | | | |
Db      21 LPANAAKL 29

RESULT 5
US-09-796-858-18
; Sequence 18, Application US/09796858
; Patent No. US20020051139A1
; GENERAL INFORMATION:
; APPLICANT: Holtzmann, Douglas
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEINS HAVING PROGNOSTIC, DIAGNOSTIC,
; FILE REFERENCE: 7853-226-999
; PREVENTIVE, THERAPEUTIC, AND OTHER USES
; CURRENT APPLICATION NUMBER: US/09/796,858
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-12
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; NUMBER OF SEQ ID NOS: 50
; SEQ ID NO 18
; LENGTH: 239
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-796-858-18
```

Query Match 62.5%: Score 30; DB 10; Length 239;
Best Local Similarity 66.7%: Pred. No. 96;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SMPLNAVK 9
| ||| |:
Db 128 SAPLNATVR 136

RESULT 6

US-09-738-626-4892
; Sequence 4892, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIRO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 4892
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4892

Query Match 62.5%: Score 30; DB 9; Length 335;
Best Local Similarity 100.0%: Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 MPLNAA 7
| ||| |:
Db 164 MPLNAA 169

RESULT 7

US-09-820-096B-5
; Sequence 5, Application US/09820096B
; Publication No. US20030022168A1
; GENERAL INFORMATION:
; APPLICANT: Kasahara, Hiroyuki
; APPLICANT: Davin, Laurence
; APPLICANT: Lewis, No. US20030022168A1man
; TITLE OF INVENTION: ARYL PROPEPAL DOUBLE BOND REDUCTASE
; FILE REFERENCE: MSUR-1-17233
; CURRENT APPLICATION NUMBER: US/09/820, 096B
; CURRENT FILING DATE: 2001-03-27
; PRIOR APPLICATION NUMBER: US 60/192, 266
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Pinus Taeda
US-09-820-096B-5

Query Match 60.4%: Score 29; DB 9; Length 18;
Best Local Similarity 55.6%: Pred. No. 8.9;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SMPLNAVK 9
| ||| |:
Db 10 SFLPNOAIR 18

RESULT 8

US-09-820-296-5
; Sequence 5, Application US/09820296
; Patent No. US20010048424A1
; GENERAL INFORMATION:
; APPLICANT: Kasahara, Hiroyuki
; APPLICANT: Davin, Laurence
; APPLICANT: Lewis, No. US20010048424A1man
; TITLE OF INVENTION: ARYL PROPEPAL DOUBLE BOND REDUCTASE
; FILE REFERENCE: MSUR-1-17233
; CURRENT APPLICATION NUMBER: US/09/820, 296
; CURRENT FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 60/192, 266
; PRIOR FILING DATE: 2000-03-27
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 18
; TYPE: PRT
; ORGANISM: Pinus Taeda
US-09-820-296-5

Query Match 60.4%: Score 29; DB 10; Length 18;
Best Local Similarity 55.6%: Pred. No. 8.9;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SMPLNAVK 9
| ||| |:
Db 10 SFLPNOAIR 18

RESULT 9

US-10-174-590-184
; Sequence 184, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174, 590
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 184
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Homo Saplen
US-10-174-590-184

Query Match 60.4%: Score 29; DB 9; Length 263;
Best Local Similarity 55.6%: Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
 |||| : : |
 Db 163 MPLNTSIVM 171

RESULT 10
 US-10-176-758-184

; Sequence 184, Application US/10176758
 ; Publication No. US20030008353A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C104
 ; CURRENT APPLICATION NUMBER: US/10/176,758
 ; FILE REFERENCE: P3430R1C104
 ; CURRENT FILING DATE: 2002-06-21
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 184
 ; LENGTH: 263
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-176-758-184

Query Match 60.4%; Score 29; DB 9; Length 263;
 Best Local Similarity 55.6%; Pred. No. 1.7e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
 |||| : : |
 Db 163 MPLNTSIVM 171

RESULT 11
 US-10-175-737-184

; Sequence 184, Application US/10175737
 ; Publication No. US20030013153A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C50
 ; CURRENT APPLICATION NUMBER: US/10/175,737
 ; CURRENT FILING DATE: 2002-06-19
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 184
 ; LENGTH: 263
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-175-737-184

Query Match 60.4%; Score 29; DB 9; Length 263;
 Best Local Similarity 55.6%; Pred. No. 1.7e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
 |||| : : |
 Db 163 MPLNTSIVM 171

RESULT 12
 US-10-173-706-184

; Sequence 184, Application US/10173706
 ; Publication No. US2003002293A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C7
 ; CURRENT APPLICATION NUMBER: US/10/173,706
 ; CURRENT FILING DATE: 2002-06-17
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 184
 ; LENGTH: 263
 ; TYPE: PRT
 ; ORGANISM: Homo Sapien
 US-10-173-706-184

Query Match 60.4%; Score 29; DB 9; Length 263;
 Best Local Similarity 55.6%; Pred. No. 1.7e+02;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 MPLNAVKM 10
 |||| : : |
 Db 163 MPLNTSIVM 171

RESULT 13
 US-10-175-738-184

; Sequence 184, Application US/10175738
 ; Publication No. US2003002294A1
 ; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C45
 ; CURRENT APPLICATION NUMBER: US/10/175,738
 ; CURRENT FILING DATE: 2002-06-19
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 184
 ; LENGTH: 263
 ; TYPE: PRT

; ORGANISM: Homo Sapien
US-10-175-738-184

Query Match
Best Local Similarity 55.6%; Score 29; DB 9; Length 263;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 MPLNAAVKM 10
Db 163 MPLNTSIWM 171

RESULT 14
US-10-175-752-184

; Sequence 184, Application US/10175752
; Publication No. US20030022295A1
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C60

; CURRENT APPLICATION NUMBER: US/10/175,752

; CURRENT FILING DATE: 2002-06-19

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 184

; LENGTH: 263

; TYPE: PRT

; ORGANISM: Homo Sapien

US-10-175-752-184

Query Match
Best Local Similarity 55.6%; Score 29; DB 9; Length 263;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 MPLNAAVKM 10
Db 163 MPLNTSIWM 171

RESULT 15
US-10-176-482-184

; Sequence 184, Application US/10176482
; Publication No. US20030022296A1
; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME

; FILE REFERENCE: P3430R1C70

; CURRENT APPLICATION NUMBER: US/10/176,482

; CURRENT FILING DATE: 2002-06-20

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 612

; SEQ ID NO 184
; LENGTH: 263
; TYPE: PRT

; ORGANISM: Homo Sapien
US-10-176-482-184

Query Match
Best Local Similarity 55.6%; Score 29; DB 9; Length 263;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 2 MPLNAAVKM 10
Db 163 MPLNTSIWM 171

Search completed: March 25, 2003, 08:23:38
Job time: 11.3939 secs

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:20:22 ; Search time 10.9091 Seconds
(without alignments)
88.123 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPLNAVKM 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|-------|----------------------------|
| 1 | 36 | 75.0 | 1009 | 2 | S49618 helicase-like tran |
| 2 | 33 | 68.8 | 459 | 2 | A99932 hypothetical prote |
| 3 | 32 | 66.7 | 154 | 2 | AF3354 translation initia |
| 4 | 32 | 66.7 | 323 | 2 | G84536 hypothetical prote |
| 5 | 32 | 66.7 | 425 | 2 | S60446 hypothetical prote |
| 6 | 32 | 66.7 | 443 | 2 | C71376 probable sodium- a |
| 7 | 32 | 66.7 | 1248 | 2 | B96827 hypothetical prote |
| 8 | 31 | 64.6 | 114 | 2 | T10673 hypothetical prote |
| 9 | 31 | 64.6 | 181 | 2 | A81899 probable amylase N |
| 10 | 31 | 64.6 | 181 | 2 | C81813 duplicated hypothe |
| 11 | 31 | 64.6 | 279 | 2 | T02471 hypothetical prote |
| 12 | 31 | 64.6 | 301 | 2 | S77169 hypothetical prote |
| 13 | 31 | 64.6 | 315 | 2 | C84096 thiorodoxin reduct |
| 14 | 31 | 64.6 | 348 | 2 | E84299 hypothetical prote |
| 15 | 31 | 64.6 | 364 | 2 | AF3363 membrane fusion pr |
| 16 | 31 | 64.6 | 421 | 2 | F72213 hypothetical prote |
| 17 | 31 | 64.6 | 425 | 2 | A83478 probable two-compo |
| 18 | 31 | 64.6 | 435 | 2 | C71807 hypothetical prote |
| 19 | 31 | 64.6 | 435 | 2 | H64693 fibronectin/fibrin |
| 20 | 31 | 64.6 | 454 | 2 | D98321 c4-dicarboxylate t |
| 21 | 31 | 64.6 | 454 | 2 | AB2962 hypothetical prote |
| 22 | 31 | 64.6 | 529 | 2 | JC5533 scavenger receptor |
| 23 | 31 | 64.6 | 650 | 2 | T04487 hypothetical prote |
| 24 | 31 | 64.6 | 666 | 2 | B69076 catalytic subunit |
| 25 | 31 | 64.6 | 823 | 2 | E64526 adenine/cytosine D |
| 26 | 31 | 64.6 | 897 | 2 | E69202 valine-tRNA ligase |
| 27 | 31 | 64.6 | 1258 | 2 | F96753 Stimilar to downy m |
| 28 | 30 | 62.5 | 47 | 2 | G91196 hypothetical prote |
| 29 | 30 | 62.5 | 71 | 2 | G81190 hypothetical prote |

| | | | | | |
|----|----|------|-----|---|---------------------------|
| 30 | 30 | 62.5 | 81 | 2 | A59368 molybdopterin bios |
| 31 | 30 | 62.5 | 113 | 2 | AD1344 hypothetical prote |
| 32 | 30 | 62.5 | 113 | 2 | AH1714 hypothetical prote |
| 33 | 30 | 62.5 | 119 | 2 | E82926 hypothetical prote |
| 34 | 30 | 62.5 | 156 | 2 | C97526 p85 protein (U6790 |
| 35 | 30 | 62.5 | 156 | 2 | AD2745 conserved hypotet |
| 36 | 30 | 62.5 | 156 | 2 | AB2717 conserved hypotet |
| 37 | 30 | 62.5 | 165 | 2 | I52634 melanin-concentrat |
| 38 | 30 | 62.5 | 181 | 2 | G81122 N-acetylmutamoyl-L |
| 39 | 30 | 62.5 | 181 | 2 | D81135 N-acetylmutamoyl-L |
| 40 | 30 | 62.5 | 189 | 2 | E97498 outer membrane ser |
| 41 | 30 | 62.5 | 269 | 2 | T16122 hypothetical prote |
| 42 | 30 | 62.5 | 288 | 2 | D83731 ABC transporter (A |
| 43 | 30 | 62.5 | 308 | 2 | F97406 ABC transporter su |
| 44 | 30 | 62.5 | 308 | 2 | AG2624 ABC transporter, m |
| 45 | 30 | 62.5 | 310 | 2 | H64497 hypothetical prote |

ALIGNMENTS

RESULT 1
S49618
helicase-like transcription factor - human
C:Species: Homo sapiens (man)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 02-Aug-2002
C:Accession: S49618; A56095
R: Ding, H.; Descheemaeker, K.; Marynen, P.; Nelles, L.; Carvalho, T.; Carmo-Fonseca, submitted to the EMBL Data Library, November 1994
A:Description: Characterization of a helicase-like transcription factor involved in p
A:Reference number: S49618
A:Accession: S49618
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1009 <DIN>
A:Cross-references: EMBL:246606; NID:9575250; PIDN:CAA86571.1; PID:9575251
R: Sheridan, P.L.; Schorpp, M.; Voz, M.L.; Jones, K.A.
J. Biol. Chem. 270, 4575-4587, 1995
A:Title: Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH m
A:Reference number: A56095; MID:95181452; PMID:7876228
A:Accession: A56095
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-34, 'P', 36-336, 'D', 338-381, 'T', 383-912, 'K', 914-1009 <SHE>
A:Cross-references: GB:L34673; NID:9531195; PIDN:AAA67436.1; PID:9531196
C:Genetics:
A:Gene: HLTF-1; GDB:SNF2L3; HTP116
A:Cross-references: GDB:392575
A:Map position: 3q25.1-3q26.1
C:Superfamily: human PML-1 protein; RING finger homology
C:Keywords: zinc
F:756-806/Domain: RING finger homology <RING>

Query Match 75.0%; Score 36; DB 2; Length 1009;
Best local Similarity 70.0%; Pred. No. 16;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPLNAVKM 10
DB 196 SMPVHAQVM 205

RESULT 2
A99932
hypothetical protein SA2117 [Imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: A99932
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Uli, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.

A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A99932
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-459 <KUR>
A:Cross-references: GB:BA000018; PID:913702126; PIDN:BA043418.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
A:Gene: SA2117

Query Match 68.8%; Score 33; DB 2; Length 459;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAAV 8
||||| 1:
Db 56 SMPLNVAI 63

RESULT 3
AF3354
Translation initiation inhibitor [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AF3354
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Ios, T.; Yanova, R.; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letessier, M.; Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3354
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-154 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAU52001.1; PID:917982764; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0820
A:Map position: 1

Query Match 66.7%; Score 32; DB 2; Length 154;
Best Local Similarity 60.0%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAAVKM 10
||||| 1:
Db 138 SLPNAPVEV 147

RESULT 4
G84536
hypothetical protein At2g16110 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: G84536
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vankken, S.E.; Unayam, L.; Tallon, L.; Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J. Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: G84536
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-323 <STO>
A:Cross-references: GB:AE002093; NID:94678207; PIDN:AA026953.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g16110
A:Map position: 2

Query Match 66.7%; Score 32; DB 2; Length 323;
Best Local Similarity 75.0%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAAVK 9
||||| 1:
Db 74 MPLNAVVR 81

RESULT 5
S60446
hypothetical protein YGR156w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein G6670
C:Species: Saccharomyces cerevisiae
C>Date: 10-Apr-1996 #sequence_revision 19-Apr-1996 #text_change 19-Apr-2002
C:Accession: S60446; S48506; S64465
R:Skala, J.; Nawrocki, A.; Goffeau, A. Yeast 11, 1421-1427, 1995
A:Title: The sequence of a 27 kb segment on the right arm of chromosome VII from Sacc

A:Reference number: S60435; MUID:96158062; PMID:8585325
A:Accession: S60446
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-425 <SKA>
A:Cross-references: EMBL:X85807; NID:91045249; PIDN:CAA59813.1; PID:91045261
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, March 1995
R:Ono, B.I.; Inoue, T.; Kijima, K.; Matsuda, A.; Negishi, K.; Shinoda, S. submitted to the EMBL Data Library, June 1993
A:Description: Identification of the structural gene of cystathionine beta-synthase 1
A:Reference number: S48505
A:Accession: S48506
A:Molecule type: DNA
A:Residues: 1-299, 'Y', 301-425 <ONO>
A:Cross-references: EMBL:D16502; NID:9391939; PIDN:BA03953.1; PID:91004469; PID:9416
R:Van Dyck, L.; Skala, J.; de Wergifosse, P.; Punelle, B.; Talla, E.; Nawrocki, A.; submitted to the Protein Sequence Database, May 1996
A:Reference number: S64428
A:Accession: S64465
A:Molecule type: DNA
A:Residues: 1-425 <VAN>
A:Cross-references: EMBL:Z79941; NID:91323264; PIDN:CAA97170.1; PID:e243711; PID:9132
A:Experimental source: strain S288C
C:Genetics:
A:Gene: SCD:PT11
A:Cross-references: SCD:S0003388
A:Map position: 7R

Query Match 66.7%; Score 32; DB 2; Length 425;
Best Local Similarity 60.0%; Pred. No. 49;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMPLNAAVKM 10
||||| 1:
Db 300 NMPLNVAWPM 309

RESULT 6
C71376
probable sodium- and chloride- dependent transporter - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: C71376
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; G rison, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; M they, L.; Weidman, J.; Smith, H.O.; Venter, J.C. Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Accession: C71376
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-443 <COL>
A:Cross-references: GB:AE001187; GB:AE000520; NID:93322273; PIDN:AA065017.1; PID:9332
A:Experimental source: strain Nichols
C:Genetics:

A:Gene: TP0023
C:Superfamily: gamma-aminobutyric acid transporter

Query Match 66.7% Score 32: DB 2: Length 443:
Best Local Similarity 50.0% Pred. No. 52:
Matches 5: Conservative 5: Mismatches 0: Indels 0: Gaps 0:

OY 1 SMPLNAVKM 10
:|||||:
Db 360 SIPLNANRV 369

RESULT 7

hypothetical protein TBK14.1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: B96827
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hultzer, L.
Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: B96827
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-1248 <STO>
A:Cross-references: GB:AE005173; NID:g4835752; PIDN:AND30219.1; GSPDB:GN00141

C:Genetics:
A:Gene: TBK14.1
A:Map position: 1

Query Match 66.7% Score 32: DB 2: Length 1248:
Best Local Similarity 60.0% Pred. No. 16+02:
Matches 6: Conservative 3: Mismatches 1: Indels 0: Gaps 0:

OY 1 SMPLNAVKM 10
:|||||:
Db 380 SIPLNAYQL 389

RESULT 8

hypothetical protein F6E21.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Oct-1999

C:Accession: T10673
R:Byvan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancro
submitted to the Protein Sequence Database, June 1999
A:Reference number: 216533

A:Accession: T10673
A:Molecule type: DNA
A:Residues: 1-114 <BEV>
A:Cross-references: EMBL:AL049914; GSPDB:GN00062; ATSP:F6E21.110

A:Experimental source: cultivar Columbia; BAC clone F6E21
C:Genetics:
A:Gene: ATSP:F6E21.110
A:Map position: 4

Query Match 64.6% Score 31: DB 2: Length 114:
Best Local Similarity 75.0% Pred. No. 20:
Matches 6: Conservative 2: Mismatches 0: Indels 0: Gaps 0:

OY 1 SMPLNAV 8
:|||||:
Db 27 SIPLNAV 34

RESULT 9

AB1899

probable amidase NMA1303 [imported] - Neisseria meningitidis (strain Z2491 serogroup

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: AB1899; E81886
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream

Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: AB1775; MUID:2022556; PMID:10761919

A:Accession: AB1899
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-181 <PAR>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84555.1; PID:g737

A:Experimental source: serogroup A, strain Z2491
A:Accession: E81886
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-181 <PA2>
A:Cross-references: GB:AL162755; GB:AL157959; NID:g7379742; PIDN:CAB84449.1; PID:g737

A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1303; NMA1188

Query Match 64.6% Score 31: DB 2: Length 181:
Best Local Similarity 60.0% Pred. No. 33:
Matches 6: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

OY 1 SMPLNAVKM 10
:|||||:
Db 57 NMPLREAVKL 66

RESULT 10

C81813

duplicated hypothetical protein NMA1864 [imported] - Neisseria meningitidis (strain Z

C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001

C:Accession: C81813
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream

Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
A:Reference number: AB1775; MUID:2022556; PMID:10761919

A:Accession: C81813
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-181 <PAR>

A:Cross-references: GB:AL162755; GB:AL157959; NID:g7380371; PIDN:CAB85087.1; PID:g738
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: NMA1864

Query Match 64.6% Score 31: DB 2: Length 181:
Best Local Similarity 60.0% Pred. No. 33:
Matches 6: Conservative 2: Mismatches 2: Indels 0: Gaps 0:

OY 1 SMPLNAVKM 10
:|||||:
Db 57 NMPLREAVKL 66

RESULT 11

T02471

hypothetical protein AC2945760 [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein F4118.26
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001

C:Accession: T02471; E84894
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; K

submitted to the EMBL Data Library, August 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F4118 genomic sequence.
 A:Reference number: Z14674
 A:Accession: T02471
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-279 <R0U>
 A:Cross-references: EMBL:AC004665; NID:g3386593; PID:g3386618
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A04420; MUID:20083487; PMID:10617197
 A:Accession: E84894
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-279 <STO>
 A:Cross-references: GB:AE002093; NID:g3386618; PID:AC28548.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: F4118.26; At2g45760
 A:Map position: 2
 A:Introns: 200/2
 C:Superfamily: Arabidopsis thaliana hypothetical protein At2g45760

Query Match 64.6%; Score 31; DB 2; Length 279;
 Best Local Similarity 62.5%; Pred. No. 52;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 MPLNAVK 9
 |||||:
 Db 64 MPINAVR 71

RESULT 12
 S77169
 hypothetical protein slr2144 - Synechocystis sp. (strain PCC 6803)
 C:Species: Synechocystis sp.
 A:Variety: PCC 6803
 C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
 C:Accession: S77169
 R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;
 O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
 DNA Res. 3, 109-136, 1996
 A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
 S.
 A:Reference number: S74322; MUID:97061201; PMID:8905231
 A:Accession: S77169
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-301 <KAN>
 A:Cross-references: EMBL:D90908; GB:AB001339; NID:g1652725; PID:BA17727.1; PID:g165280
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
 C:Genetics:
 A:Start codon: GTG
 C:Superfamily: Synechocystis hypothetical protein slr2144

Query Match 64.6%; Score 31; DB 2; Length 301;
 Best Local Similarity 85.7%; Pred. No. 57;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPLNAA 7
 |:|||||
 Db 65 SLPLNAA 71

RESULT 13
 C84096
 thioredoxin reductase (NADPH) trxB [imported] - Bacillus halodurans (strain C-125)
 C:Species: Bacillus halodurans
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: C84096

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: C84096
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-315 <STO>
 A:Cross-references: GB:AF001519; GB:BA000004; NID:g10176109; PID:BA807290.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: trxB
 C:Superfamily: thioredoxin reductase; thioredoxin reductase homology

Query Match 64.6%; Score 31; DB 2; Length 315;
 Best Local Similarity 75.0%; Pred. No. 60;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVK 9
 |||||
 Db 243 LPLNAVK 250

RESULT 14
 E84299
 hypothetical protein Vng1455h [imported] - Halobacterium sp. NRC-1
 C:Species: Halobacterium sp. NRC-1
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: E84299
 R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Bergquist, B.; Pan, M.; Shukla, H.D.; Lasky
 ; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
 Jung, K.H.; Alam, M.; Freitas, T.
 Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
 A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.;
 A:Title: Genome sequence of Halobacterium species NRC-1.
 A:Reference number: A84160; MUID:20504483; PMID:11016950
 A:Accession: E84299
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-348 <STO>
 A:Cross-references: GB:AE004437; NID:g10580951; PID:AG19761.1; GSPDB:GN00138
 C:Genetics:
 A:Gene: VNG1455H

Query Match 64.6%; Score 31; DB 2; Length 348;
 Best Local Similarity 75.0%; Pred. No. 66;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAAV 8
 |:|||||
 Db 33 STPLNAAI 40

RESULT 15
 AF3363
 membrane fusion protein mtrC [imported] - Brucella melitensis (strain 16M)
 C:Species: Brucella melitensis
 C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
 C:Accession: AF3363
 R:DelVecchio, V.G.; Kaparat, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
 ; Mazur, M.; Goltsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Let
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
 A:Reference number: AD3252; PMID:11756688
 A:Accession: AF3363
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-364 <KUR>
 A:Cross-references: GB:AE008917; PID:AA152073.1; PID:g17982843; GSPDB:GN00190
 A:Experimental source: strain 16M
 C:Genetics:
 A:Gene: BMEI0892
 A:Map position: 1

Query Match 64.6%; Score 31; DB 2; Length 364;
 Best Local Similarity 60.0%; Pred. No. 70;
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 SMPLNAAVKM 10
 :||| :
 Db 126 AMPANAAVAL 135
 Search completed: March 25, 2003, 08:22:16
 Job time : 12.9091 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:18:08 ; Search time 6.0661 Seconds
(without alignments)
68.436 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPUNAAYKM 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt.40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match Length | ID | Description |
|------------|-------|--------------------|-------------------|--------------------|
| 1 | 32 | 66.7 | 425 1 YG30_YEAST | P39927 saccharomyc |
| 2 | 31 | 64.6 | 877 1 SYV_METH | O28861 methanobact |
| 3 | 31 | 64.6 | 1522 1 MRP3_RAT | O88563 rattus nov |
| 4 | 30 | 62.5 | 81 1 MOAD_HAERIN | P45309 haemophilus |
| 5 | 30 | 62.5 | 119 1 Y167_UREPA | O9696 ureaplasma |
| 6 | 30 | 62.5 | 165 1 MICR_HUMAN | P20382 homo sapien |
| 7 | 30 | 62.5 | 310 1 YP85_METYA | O59980 methanococ |
| 8 | 30 | 62.5 | 423 1 GCH2_CORAM | O24752 corynebacte |
| 9 | 30 | 62.5 | 456 1 YP95_YEAST | O02891 saccharomyc |
| 10 | 30 | 62.5 | 805 1 PIF1_SCHPO | O9uua2 schizosacch |
| 11 | 30 | 62.5 | 810 1 PLSB_HAERIN | P44857 haemophilus |
| 12 | 30 | 62.5 | 830 1 FAR1_YEAST | P21268 saccharomyc |
| 13 | 30 | 62.5 | 1020 1 ACAL_ARATH | O37145 arabidopsis |
| 14 | 30 | 62.5 | 1134 1 TIE1_MOUSE | O06806 mus musculu |
| 15 | 30 | 62.5 | 1136 1 TIE1_BOVIN | O06805 bos taurus |
| 16 | 30 | 62.5 | 1138 1 TIE1_HUMAN | P35590 homo sapien |
| 17 | 30 | 62.5 | 1839 1 CYAA_SACKL | P23466 saccharomyc |
| 18 | 29 | 60.4 | 128 1 Y082_CAEEL | O09237 caenorhabdi |
| 19 | 29 | 60.4 | 263 1 TMAA_HUMAN | O43736 homo sapien |
| 20 | 29 | 60.4 | 263 1 TMAA_HUMAN | O43736 homo sapien |
| 21 | 29 | 60.4 | 269 1 MHPD_MOUSE | O61500 mus musculu |
| 22 | 29 | 60.4 | 336 1 USG_PSEAE | P77608 escherichia |
| 23 | 29 | 60.4 | 344 1 DCUP_DEIRA | O87014 pseudomonas |
| 24 | 29 | 60.4 | 363 1 Y10N_BACSU | O91v96 deinococcus |
| 25 | 29 | 60.4 | 425 1 CABL_CAEEL | O34617 bacillus su |
| 26 | 29 | 60.4 | 447 1 GNT2_HUMAN | O93249 caenorhabdi |
| 27 | 29 | 60.4 | 470 1 SYC_SULSO | O10469 h alpha-1,6 |
| 28 | 29 | 60.4 | 620 1 YJB9_YEAST | O97w66 sulfolobus |
| 29 | 29 | 60.4 | 668 1 FEOB_METYA | P47069 saccharomyc |
| 30 | 29 | 60.4 | 693 1 YERT_SCHPO | O57986 methanococ |
| 31 | 29 | 60.4 | 696 1 SPOT_AQUAE | O14286 schizosacch |
| 32 | 29 | 60.4 | 941 1 GCSP_MYCTU | O67012 aquifex aeo |
| 33 | 29 | 60.4 | 950 1 MCI1_YEAST | O50601 mycobacteri |
| | | | | P53258 saccharomyc |

| | | | | |
|----|----|------|-------------------|---------------------|
| 34 | 29 | 60.4 | 956 1 GCSP_ECOLI | P33195 escherichia |
| 35 | 29 | 60.4 | 1070 1 PVDC_PLANK | P50494 plasmodium |
| 36 | 29 | 60.4 | 1790 1 VIT_ANTGR | O05808 anthomomus |
| 37 | 29 | 60.4 | 2144 1 BP28_HUMAN | O9h83 homo sapien |
| 38 | 29 | 60.4 | 4870 1 RYR3_HUMAN | O15413 homo sapien |
| 39 | 28 | 58.3 | 110 1 Y250_ARCFU | O29989 archaeoglob |
| 40 | 28 | 58.3 | 125 1 Y364_AQUAE | O66889 aquifex aeo |
| 41 | 28 | 58.3 | 145 1 MPE1_YEAST | P40185 saccharomyc |
| 42 | 28 | 58.3 | 187 1 KAD_XYLF | O99gm3 xyloella fas |
| 43 | 28 | 58.3 | 236 1 FRE_VIBFI | P43126 vibrio fisc |
| 44 | 28 | 58.3 | 245 1 NCAP_RVFW | P21700 rift valley |
| 45 | 28 | 58.3 | 258 1 CTR2_ANOGA | O17025 anopheles g |

ALIGNMENTS

RESULT 1
ID YG30_YEAST STANDARD: PRT: 425 AA.
AC P39927;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1996 (rel. 33, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Hypothetical 47.0 kDa protein in CYS4-PEM1 intergenic region.
GN YGRI56W OR G6670.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=45-8-1A;
RA Ono B.I., Inoue T., Kijima K., Matsuda A., Negishi K., Shinoda S.;
RL Submitted (JUN-1993) to the EMBL/GenBank/DBD databases.
RM [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=96158062; PubMed=8585325;
RA Skala J., Nawrocki A., Goffeau A.;
RT "The sequence of a 27 kb segment on the right arm of chromosome VII
from Saccharomyces cerevisiae reveals MOLI, NAT2, RPL30B, RSR1, CYS4,
PEM1/CHO2, NSR1 genes and ten new open reading frames.";
RL Yeast 11:1421-1427(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).
CC -----
CC
DR EMBL: D16502; BAA03953.1; -;
DR EMBL: X85807; CAA59813.1; -;
DR EMBL: Z72941; CAA97170.1; -;
DR PIR: S48506; S48506.
DR SGD: S0003388; YGR156W.
DR InterPro: IPR000504; RNA_rec_mot.
DR SMART: SM00360; RRM; 1.
KW Hypothetical protein.
FT CONFLICT 300 300
SQ SEQUENCE 425 AA; 46983 MW; F6564311A2B1F240 CRC64;
N -> Y (IN REF. 1).

Query Match 66.7%; Score 32; DB 1; Length 425;

Best Local Similarity 60.0%; Pred. No. 20;
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 SMPUNAAYKM 10
DB 300 NMPUNAYMPP 309

```

RESULT 2
SYV_METTH STANDARD; PRT; 877 AA.
AC 026861;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Valyl-tRNA synthetase (EC 6.1.1.9) (Valine--tRNA ligase) (VALRS).
GN VALS OR MTH67.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-Delta H;
RX MEDLINE=98037514; PubMed=9371463;
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Kesgle P., Lumm W., Pothler B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
CC -1- CATALYTIC ACTIVITY: ATP + L-valine + tRNA(Val) = AMP + diphosphate
CC + L-valyl-tRNA(Val).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000855; AAB85270.1; -.
DR HSSP; P96142; IGAX.
DR InterPro: IPR002300; tRNA-synt_1a.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR002303; tRNA-synt_val.
DR Pfam: PF00133; tRNA-synt_1; 1.
DR PRINTS; PR00986; TRNASYNTHAL.
DR TIGRfams; TIGR00422; VALS; 1.
DR PROSITE; PS00178; AA-TRNA-LIGASE_I; 1.
KM Aminoacyl-tRNA synthetase; Protein biosynthesis; ligase; ATP-binding;
KM Complete proteome.
FT SITE 46 "HIGH" REGION.
FT SITE 529 "KMSKS" REGION.
FT BINDING 532 532 ATP (BY SIMILARITY).
SQ SEQUENCE 877 AA; 100946 MW; B939720D6DF0DB CRC64;

Query Match 64.68; Score 31; DB 1; Length 877;
Best Local Similarity 75.08; Pred. No. 72;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

DE resistance-associated protein 3) (MRP-like protein-2) (MLP-2).
GN ABCB3 OR CMOAT2 OR MRP3 OR MLP2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Liver;
RX MEDLINE=99292429; PubMed=10362653;
RA Ortiz D.F., Li S., Iyer R., Zhang X., Novikoff P., Atlas I.M.;
RT "MRP3, a new ATP-binding cassette protein localized to the canalicular
RT domain of the hepatocyte.";
RL Am. J. Physiol. 276:G1493-G1500(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=Sprague-Dawley; TISSUE=Colon;
RX MEDLINE=98279126; PubMed=9614210;
RA Hltoshashi T., Suzuki H., Ito K., Ogawa K., Kume K., Shinazu T.,
RA Sugiyama Y.;
RT "Hepatic expression of multidrug resistance-associated protein-like
RT proteins maintained in elsal hyperbilirubinemic rats and TR(-) mutant rats."
RL Mol. Pharmacol. 53:1068-1075(1998).
CC -1- FUNCTION: MAY ACT AS AN INDUCIBLE TRANSPORTER IN THE BILIARY AND
CC INTESTINAL EXCRETION OF ORGANIC ANIONS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: LUNG AND INTESTINE, LOW IN LIVER. HIGHER IN
CC LIVER OF ELSAI HYPERBILIRUBINEMIC RATS AND TR(-) MUTANT RATS.
CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF072816; AAC25416.1; -.
DR EMBL; AB010467; BAA28955.1; -.
DR HSSP; P13569; INBD.
DR InterPro: IPR003593; AAA ATPase.
DR InterPro: IPR003439; ABC transporter.
DR InterPro: IPR001140; ABCtransprtm.
DR InterPro: IPR005292; MRP_assoc.
DR Pfam; PF00005; ABC_tran; 2.
DR Pfam; PF00664; ABC_membrane; 2.
DR ProDom; PD000006; ABC_transportr; 2.
DR SMART; SM00382; AAA; 2.
DR TIGRfams; TIGR00957; MRP_assoc_pro; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
FT ATP-binding; Glycoprotein; Transmembrane; Transport; Repeat.
FT DOMAIN 1 32 EXTRACELLULAR (BY SIMILARITY).
FT TRANSSEM 33 53 1 (BY SIMILARITY).
FT DOMAIN 54 73 CYTOPLASMIC (BY SIMILARITY).
FT TRANSSEM 74 94 2 (BY SIMILARITY).
FT DOMAIN 95 99 EXTRACELLULAR (BY SIMILARITY).
FT TRANSSEM 100 120 3 (BY SIMILARITY).
FT DOMAIN 121 132 CYTOPLASMIC (BY SIMILARITY).
FT TRANSSEM 133 153 4 (BY SIMILARITY).
FT DOMAIN 154 171 EXTRACELLULAR (BY SIMILARITY).
FT TRANSSEM 172 192 5 (BY SIMILARITY).
FT DOMAIN 193 301 CYTOPLASMIC (BY SIMILARITY).
FT TRANSSEM 302 322 6 (BY SIMILARITY).
FT DOMAIN 323 347 EXTRACELLULAR (BY SIMILARITY).
FT TRANSSEM 348 368 7 (BY SIMILARITY).
FT DOMAIN 369 424 CYTOPLASMIC (BY SIMILARITY).
FT TRANSSEM 425 445 8 (BY SIMILARITY).
FT DOMAIN 446 448 EXTRACELLULAR (BY SIMILARITY).
FT TRANSSEM 449 469 9 (BY SIMILARITY).
FT DOMAIN 470 531 CYTOPLASMIC (BY SIMILARITY).
FT TRANSSEM 532 552 10 (BY SIMILARITY).
FT DOMAIN 553 574 EXTRACELLULAR (BY SIMILARITY).

```

```

FT TRANSNM 575 595 11 (BY SIMILARITY).
FT DOMAIN 596 958 CYTOPLASMIC (BY SIMILARITY).
FT TRANSNM 959 979 12 (BY SIMILARITY).
FT DOMAIN 980 1016 EXTRACELLULAR (BY SIMILARITY).
FT TRANSNM 1017 1037 13 (BY SIMILARITY).
FT DOMAIN 1038 1080 CYTOPLASMIC (BY SIMILARITY).
FT TRANSNM 1081 1101 14 (BY SIMILARITY).
FT DOMAIN 1102 1102 EXTRACELLULAR (BY SIMILARITY).
FT TRANSNM 1103 1123 15 (BY SIMILARITY).
FT DOMAIN 1124 1194 CYTOPLASMIC (BY SIMILARITY).
FT TRANSNM 1195 1215 16 (BY SIMILARITY).
FT DOMAIN 1216 1217 EXTRACELLULAR (BY SIMILARITY).
FT TRANSNM 1218 1238 17 (BY SIMILARITY).
FT DOMAIN 1239 1522 CYTOPLASMIC (BY SIMILARITY).
FT NP_BIND 659 666 ATP (POTENTIAL).
FT NP_BIND 1318 1325 ATP (POTENTIAL).
FT CARBOHYD 1001 1001 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1002 1002 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 323 344 SPSTHSCASASSGELFPPHCPY -> LSFINPOLLSILIRF
FT CONFLICT 645 645 ISDPTAPT (IN REF. 2).
FT CONFLICT 1075 1075 I -> L (IN REF. 2).
FT CONFLICT 1075 1075 H -> D (IN REF. 2).
SQ SEQUENCE 1522 AA; 168977 MW; 740E31E0C4C64297 CRC64;

Query Match 64.6%; Score 31; DB 1; Length 1522;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 MPLNAVKM 10
Db 461 IPLNGAVSM 469

RESULT 4
MOAD_HAEIN STANDARD; PRT; 81 AA.
ID MOAD_HAEIN
AC P45309;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Molybdopterin converting factor subunit 1 (MPT synthase subunit 1)
DE (Molybdopterin synthase subunit 1) (Molybdenum cofactor biosynthesis
protein D) (Molybdopterin converting factor small subunit).
GN MOAD OR H11674.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN NCBI_TaxID=727;
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Frithman J.L., Geoghagen N.S.M.,
RA Grehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.:
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Converts molybdopterin precursor 2 into molybdopterin.
CC This requires the incorporation of two sulfur atoms into precursor
CC 2 to generate a dithiolene group. The sulfur atoms are provided by
CC the active form of the small subunit (By similarity).
```

```

CC -!- PATHWAY: Molybdenum cofactor biosynthesis.
CC -!- SUBUNIT: Heterodimer of 2 moad subunits and 2 moae subunits (By
CC similarity).
CC -!- SIMILARITY: BELONGS TO THE MOAD FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32840; AAC2319.1; -.
DR HSSP: P30748; IFMO.
DR TIGR: H11674; -.
DR InterPro: IPR003749; THIS.
DR Pfam: PF02597; DUF170; 1.
KW Molybdenum cofactor biosynthesis; Complete proteome.
SQ SEQUENCE 81 AA; 8826 MW; 35D140F82456F22 CRC64;
```

```

OY 2 MPLNAVK 9
Db 60 MPLESAVK 67

Query Match 62.5%; Score 30; DB 1; Length 81;
Best Local Similarity 75.0%; Pred. No. 9.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

RESULT 5
Y167_UREPA STANDARD; PRT; 119 AA.
ID Y167_UREPA
AC Q9POX6;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein U0167.
GN U0167.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Ureaplasma.
OC NCBI_TaxID=134821;
RN NCBI_TaxID=134821;
RP SEQUENCE FROM N.A.
RC STRAIN=serovar 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum";
RL Nature 407:757-762(2000).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE002116; AAF30574.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 119 AA; 14280 MW; F58A475E92C8BP9 CRC64;
```

| RESULT 6 | MLCH_HUMAN | STANDARD: | PRT: 165 AA. |
|----------|--|-----------|--------------|
| ID | MICH_HUMAN | | |
| AC | P20382; Q16044; | | |
| DT | 01-FEB-1991 (Rel. 17, Created) | | |
| DT | 15-JUL-1999 (Rel. 38, Last sequence update) | | |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) | | |
| DE | PRO-MCH precursor [Contains: Neuropeptide-glycine-glutamic acid (NGE)] | | |
| DE | (Neuropeptide G-E); Neuropeptide-glutamic acid-isoleucine (NEI) | | |
| DE | (Neuropeptide E-1); Melanin-concentrating hormone (MCH)]. | | |
| GN | PMCH OR MCH. | | |
| OS | Homo sapiens (Human). | | |
| OC | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; | | |
| OC | Mammalia; Eutheria; Primates; Carnivora; Hominidae; Homo. | | |
| OX | NCBI_TaxID:9606; | | |
| RN | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RC | TISSUE-Hypothalamus; | | |
| RX | MEDLINE-91125371; PubMed-2149166; | | |
| RA | Presse F., Nahon J.-L., Fischer W.H., Vale W.; | | |
| RT | "Structure of the human melanin concentrating hormone mRNA."; | | |
| RT | Mol. Endocrinol. 4:632-637(1990). | | |
| RL | [2] | | |
| RN | SEQUENCE FROM N.A. | | |
| RP | TISSUE-Breast cancer; | | |
| RC | MEDLINE-93316802; PubMed-8326825; | | |
| RX | Breton C., Schorpp M., Nahon J.-L.; | | |
| RA | "Isolation and characterization of the human melanin-concentrating | | |
| RT | hormone gene and a variant gene."; | | |
| RL | Brain Res. Mol. Brain Res. 18:297-310(1993). | | |
| RN | [3] | | |
| RP | PROCESSING. | | |
| RX | MEDLINE-99156937; PubMed-10037747; | | |
| RA | Vale A., Ortolia C., Hervieu G., Furuta M., Barbero P., Steiner D.F., | | |
| RT | Seidah N.G., Nahon J.-L.; | | |
| RT | "Cellular localization and role of pro-melanin-concentrating hormone in the | | |
| RL | processing of pro-melanin concentrating hormone in mammals."; | | |
| RN | J. Biol. Chem. 274:6536-6545(1999). | | |
| RP | [4] | | |
| RX | TISSUE SPECIFICITY. | | |
| RA | MEDLINE-97334402; PubMed-9191099; | | |
| RT | Vale A., Zhixing Y., Breton C., Pedetour F., Coquerel A., Jordan D., | | |
| RT | Nahon J.-L.; | | |
| RT | "The melanin-concentrating hormone gene in human: flanking region | | |
| RL | analysis, fine chromosome mapping, and tissue-specific expression."; | | |
| CC | Brain Res. Mol. Brain Res. 46:243-255(1997). | | |
| CC | -1- FUNCTION: MCH may act as a neurotransmitter or neuromodulator in a | | |
| CC | broad array of neuronal functions directed toward the regulation | | |
| CC | of goal-directed behavior, such as food intake, and general | | |
| CC | arousal. May also have a role in spermatocyte differentiation. | | |
| CC | -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN LATERAL | | |
| CC | HYPOTHALAMUS. ALSO DETECTED IN PALIDIDIUM, NEOCORTEX AND | | |
| CC | CEREBELLUM. ALSO FOUND IN THYMUS, BROWN ADIPOSE TISSUE, DUODENUM | | |
| CC | AND TESTIS (SPERMATOGENIA, EARLY SPERMATOCTES AND SERTOLI CELLS). | | |
| CC | NO EXPRESSION IN PERIPHERAL BLOOD. IN BRAIN EXCLUSIVELY MATURE MCH | | |
| CC | AND NEI PEPTIDES ARE PRESENT. IN PERIPHERAL TISSUES A LARGE | | |
| CC | PRODUCT, ENCOMPASSING THE NEI AND MCH DOMAINS OF THE PRECURSOR, IS | | |
| CC | FOUND PREDOMINANTLY. | | |
| CC | -1- PTM: DIFFERENTIALLY PROCESSED IN THE BRAIN AND IN PERIPHERAL | | |
| CC | ORGANS PRODUCING TWO NEUROPEPTIDES; NEI AND MCH. A THIRD PEPTIDE, | | |
| CC | NGE, MAY ALSO BE PRODUCED. PREFERENTIAL PROCESSING IN NEURONS BY | | |
| CC | PROHORMONE CONVERTASE 2 (PC2) GENERATES NEI. MCH IS GENERATED IN | | |
| CC | NEURONS OF THE LATERAL HYPOTHALAMIC AREA BY SEVERAL PROHORMONE | | |
| CC | CONVERTASES INCLUDING PC1/3, PC2 AND PC5/6. | | |
| CC | -1- PTM: MCH is a cyclic peptide. | | |
| CC | -1- SIMILARITY: BELONGS TO THE MELANIN-CONCENTRATING HORMONE FAMILY. | | |
| CC | ----- | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration | | |
| CC | between the Swiss Institute of Bioinformatics and the EMBL outstation - | | |
| CC | the European Bioinformatics Institute. There are no restrictions on its | | |
| CC | use by non-profit institutions as long as its content is in no way | | |
| CC | modified and this statement is not removed. Usage by and for commercial | | |
| CC | entities requires a license agreement (See http://www.isb-sib.ch/announce/) | | |

```

CC ----- send an email to license@lsb-sib.ch). -----
CC
DR EMBL: M57703; AAA63214.1; -.
DR EMBL: S63697; AB827493.1; -.
DR PIR: A34551; A34551.
DR Genew: HGNC:9109; PMCH.
DR MIM: 176795; -.
KW Cleavage on pair of basic residues; Hormone; Neuropeptide; Signal;
KW Amidation; Spermatogenesis.
FT SIGNAL 1 21
FT CHAIN 22 165
FT PEPTIDE 110 128
FT PEPTIDE 131 143
FT PEPTIDE 147 165
FT MOD_RES 143 143
FT DISULFID 153 162
FT CONFLICT 42 42
FT CONFLICT 104 106
FT CONFLICT 113 113
FT SEQUENCE 165 AA; 18723 MW; D639E8388637244 CRC64;
SQ
Query Match 62.5%; Score 30; DB 1; Length 165;
Best Local Similarity 62.5%; Pred. NO. 20;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MFLNAVK 9
Db 96 LPLMLAIK 103
:::|:|:|
RESULT 7
YF85_METJA STANDARD: PRT; 310 AA.
ID YF85_METJA
AC O58980;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative aldolase MJ1585 (EC 4.2.1.-).
GN MJ1585.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=DAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; Pubmed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kellum J.A., Blake J.A., Fitch J.F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weissbrock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Ullrich J., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE DEOC/FBAB FAMILY OF ALDOLASES.
CC -----
CC CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
CC EMBL: U67598; AAB99604.1; -.
CC TIGR: MJ1585; -.
CC InterPro: IPR002915; Deoc.
CC Pfam: PF01791; Deoc; 1.

```


KW Hypothetical protein; Lyase; Schiff base; Complete proteome.
 FT BINDING 213 213 SCHIFF-BASE (BY SIMILARITY).
 SQ SEQUENCE 310 AA; 34574 MW; 4AB6286F888049C0 CRC64;

Query Match 62.5%; Score 30; DB 1; Length 310;
 Best Local Similarity 62.5%; Pred. No. 39;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 PLNAAVKM 10
 DB 277 PLDAAIRM 284

RESULT 8

GCH2_CORAM STANDARD; PRT; 423 AA.

ID GCH2_CORAM
 AC 024752; 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Riboflavin biosynthesis protein ribA (includes: GTP cyclohydrolase II
 DE (EC 3.5.4.25); 3,4-dihydroxy-2-butanone 4-phosphate synthase (DHBP
 DE synthase)).
 GN RIBA.

OS Corynebacterium ammoniagenes (Brevibacterium ammoniagenes).
 OC Bacteria; Actinobacteria; Actinobacteriales (class); Actinobacteridae;
 OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
 OC Corynebacterium.

NCBI_TaxID=1697;
 RN [1]

RP SEQUENCE FROM N.A.
 RA Koizumi S., Yonetani Y., Teshiba S.;
 RT "Process for producing riboflavin.";
 RL Patent number US5589355, 31-DEC-1996.

CC -1- CATALYTIC ACTIVITY: GTP + 3 H₂O = formate + 2,5-diamino-6-
 CC hydroxy-4-(5-phosphoribosylamino)pyrimidine + diphosphate.
 CC -1- PATHWAY: Riboflavin biosynthesis.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE DHBP
 CC SYNTHASE FAMILY.

CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE GTP
 CC CYCLOHYDROLASE II FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 CC EMBL: AB003693; BAA20055.1; -
 CC InterPro: IPR000422; DHBP_synthase.
 CC InterPro: IPR000926; GTP_cyclohydrol2.
 CC Pfam: PF00925; GTP_cyclohydrol2; 1.
 CC DR Pfam: PF00926; DHBP_synthase; 1.
 CC DR Pfam: PD003034; DHBP_synthase; 1.
 CC DR TIGRfams: TIGR00505; ribA; 1.
 CC DR TIGRfams: TIGR00506; ribB; 1.
 CC KW Multifunctional enzyme; Riboflavin biosynthesis; Hydrolase.

FT DOMAIN 1 210 DHBP SYNTHASE.
 FT DOMAIN 211 423 GTP CYCLOHYDROLASE II.
 SQ SEQUENCE 423 AA; 46399 MW; EBBE9CA3E902BD25 CRC64;

Query Match 62.5%; Score 30; DB 1; Length 423;
 Best Local Similarity 62.5%; Pred. No. 55;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 3 PLNAAVKM 10
 DB 4 PLNSAVRL 11

RESULT 9

YP95_YEAST STANDARD; PRT; 456 AA.
 ID YP95_YEAST
 AC 002891;

DT 01-NOV-1997 (rel. 35, Created)
 DT 01-NOV-1997 (rel. 35, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Hypothetical protein YPL095C.
 DE YPL095C OR Lpg13C.
 GN

OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

NCBI_TaxID=4932;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RX MEDLINE=97313271; PubMed=9169875;

RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansoorge W.,
 RA Araujo R., Aparicio A., Barrell B.G., Badcock K., Benes V.,
 RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
 RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
 RA Dietrich F.S., Dilius H., DiPaolo T., Dubois E., Duesterhoeft A.,
 RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
 RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.,
 RA Hunkeler-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
 RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
 RA Marathe R., Messenguy F., Mewes H.-W., Milpail S., Moestl D.,
 RA Mueller-Auer S., Namath A., Nentwich U., Oelfner P., Pearson D.,
 RA Petel F.X., Pohl T.M., Purnelle D., Schafer M., Scharte M.,
 RA Scherens B., Schramm S., Schroeder M., Sidcu A.W., Tetteil H.,
 RA Ureastarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
 RA Walsh S.V., Wambutt R., Wang Y., Wedler H., Wedler H., Minnett E.,
 RA Zhong W.W., Zollner A., Vo D.H., Hanl J.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
 RL Nature 387:103-105(1997).
 CC -1- SIMILARITY: BELONGS TO THE UPF0017 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

CC -----
 CC EMBL: U43281; AAB68204.1; -
 CC SCD: S0006016; YPL095C.
 CC InterPro: IPR000073; Abhydrolase.
 CC InterPro: IPR000379; Ser_estrs_site.
 CC InterPro: IPR000952; UPF0017.
 CC Pfam: PF00561; abhydrolase; 1.
 CC DR PROSITE: PS01133; UPF0017; 1.

KW Hypothetical protein
 SQ SEQUENCE 456 AA; 51723 MW; CAB660F44A287E CRC64;

Query Match 62.5%; Score 30; DB 1; Length 456;
 Best Local Similarity 75.0%; Pred. No. 60;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 3 PLNAAVKM 10
 DB 269 PLNAAVAL 276

RESULT 10

PIF1_SCHPO STANDARD; PRT; 805 AA.
 ID PIF1_SCHPO
 AC 090UA2; 074239;

DT 16-OCT-2001 (rel. 40, Created)
 DT 16-OCT-2001 (rel. 40, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)

DE DNA repair and recombination protein pif1, mitochondrial precursor.
 GN PIF1 OR PFI1 OR SPBC887.14C.
 OS Schizosaccharomyces pombe (fission yeast).

OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_TaxID=4896;
 RN [1]
 RN SEQUENCE FROM N.A.
 RA Zhou J.-Q., Qi H., Schulz V.P., Mateyak M.K., Monson E.K.,
 RA Zakian V.A.;
 RT "Schizosaccharomyces pombe pif1+ encodes an essential 5' to 3' DNA
 RT helicase that is a member of the Pif1 sub-family of DNA helicases.";
 RL Mol. Biol. Cell 0:0-0(2002).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Miblett D., Odell C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Voicakert G., Aert R., Robben J., Grymoprez B.,
 RA Weljans I., Vastreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 RA Eger B., Zimmermann W., Weiler H., Wambolt R., Punnelle B.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Rhode J.,
 RA Dacia R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Sipakovski G.V., Usery D., Bartell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe.";
 RL Nature 415:871-880(2002).
 CC -1- FUNCTION: Required for both repair of mitochondrial DNA and
 CC recognition of a recombinogenic signal characterized by a 26-bp
 CC palindromic at sequence in the very region of mitochondrial DNA.
 CC This is a single stranded DNA-dependent ATPase and DNA helicase
 CC which unwinds partially duplex DNA in a 5' to 3' polarity.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. RAD3/XPD SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF074944; AAC26139.1; -;
 DR EMBL: AL033388; CAA21899.1; -;
 KM Mitochondrion; ATP-binding; DNA-binding; DNA repair;
 KM DNA recombination; Helicase; Transist peptide.
 FT TRANSIT 1
 FT CHAIN ? 805
 FT PIF1.
 FT NP_BIND 332 339
 FT DNA_BIND 744 763
 FT CONFLICT 203 203 T -> A (IN REF. 1).
 SQ SEQUENCE 805 AA; 90049 MW; AD6DA44B47C8CABP CRC64;
 Query Match 62.5%; Score 30; DB 1; Length 805;
 Best Local Similarity 66.7%; Pred. No. 1.le+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAAVK 9
 ID PLSB_HAEIN STANDARD; PRT; 810 AA.
 AC P44857;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycerol-3-phosphate acyltransferase (EC 2.3.1.15) (GPAT).
 GN PLSB OR H10748.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 CC Haemophilus
 OX NCBI_TaxID=727;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Rd / KW20 / ATCC 51907;
 RC MEDLINE=95350630; PubMed=7542800;
 RX Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kesteven A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McEwen K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fife L.D., Fitchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd.";
 RL Science 269:496-512(1995).
 CC -1- CATALYTIC ACTIVITY: Acyl-CoA + sn-glycerol 3-phosphate = CoA + 1-
 CC acyl-sn-glycerol 3-phosphate.
 CC -1- PATHWAY: DE NOVO PHOSPHOLIPID BIOSYNTHESIS; FIRST STEP. MAY ALSO
 CC FUNCTION IN THE REGULATION OF MEMBRANE BIOGENESIS.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE GPAT / DAPAT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: U32758; AAC22406.1; -;
 DR TIGR: HI0748; -;
 DR InterPro: IPR002123; Acyltransferase.
 DR Pfam: PF01553; Acyltransferase; 1.
 KM Phospholipid biosynthesis; Transferase; Acyltransferase; Membrane;
 KM Complete proteome.
 SQ SEQUENCE 810 AA; 92774 MW; DB8564BC3E2C370D CRC64;
 Query Match 62.5%; Score 30; DB 1; Length 810;
 Best Local Similarity 66.7%; Pred. No. 1.le+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 OY 1 SMPLNAAVK 9
 ID PLSB_HAEIN STANDARD; PRT; 810 AA.
 AC P21268;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Cyclin-dependent kinase inhibitor FAR1 (CKI FAR1) (Factor arrest
DE protein).
GN FAR1 OR YUL157C OR J0565.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91077938; PubMed=2147873;
RA Chang F., Herskowitz I.;
RT "Identification of a gene necessary for cell cycle arrest by a
RT negative growth factor of yeast: FAR1 is an inhibitor of a G1 cyclin,
RT Cln2.";
RL Cln2 63:999-1011.(1990).
RN [2]
RP REVISIONS TO N-TERMINUS.
RX MEDLINE=95257932; PubMed=7739534;
RA McKinney J.D., Cross F.R.;
RT "FAR1 and the G1 phase specificity of cell cycle arrest by mating
RT factor in Saccharomyces cerevisiae.";
RL Mol. Cell. Biol. 15:2509-2516(1995).
RN [3]
RP SEQUENCE FROM N.A.
RA Obermaier B., Piravandi E., Rinke M., Domdey H.;
RN Submitted (SEP-1995) to the EMBL/Genbank/DBJ databases.
RL [4]
RP REVIEW.
RA Chang F.;
RT "Stop that cell cycle.";
RL Curr. Biol. 3:693-695(1993).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=94345395; PubMed=8066461;
RA Peter M., Herskowitz I.;
RT "Direct inhibition of the yeast cyclin-dependent kinase Cdc28-Cln by
RT Far1.";
RL Science 265:1228-1231(1994).
RN [6]
RP CHARACTERIZATION.
RX MEDLINE=96074755; PubMed=7490290;
RA Valtz N., Peter M., Herskowitz I.;
RT "FAR1 is required for oriented polarization of yeast cells in
RT response to mating pheromones.";
RL J. Cell Biol. 131:863-873(1995).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE=97394705; PubMed=9251033;
RA Valtz N., Peter M.;
RT "Functional analysis of FAR1 in yeast.";
RL Meth. Enzymol. 283:350-365(1997).
RN [8]
RP CHARACTERIZATION.
RX MEDLINE=98037745; PubMed=9367986;
RA Henchoz S., Chl Y., Catarin B., Herskowitz I., Deshaies R.J.,
RA Peter M.;
RT "Phosphorylation and ubiquitin-dependent degradation of the cyclin-
RT dependent kinase inhibitor Far1p in budding yeast.";
RL Genes Dev. 11:3046-3060(1997).
RN [9]
RP MUTAGENESIS OF SER-87 AND THR-306.
RX MEDLINE=98298210; PubMed=9632750;
RA Gartner A., Jovanovic A., Jeoung D.I., Bourlat S., Cross F.R.,
RA Amerer G.;
RT "Pheromone-dependent G1 cell cycle arrest requires Far1
RT phosphorylation, but may not involve inhibition of Cdc28-Cln2 kinase,
RT in vivo.";
RL Mol. Cell. Biol. 18:3681-3691(1998).
CC -1- FUNCTION: INHIBITOR OF THE CYCLIN-DEPENDENT KINASE CDC28.
CC CONTRIBUTES TO CELL CYCLE ARREST. INVOLVED IN PHEROMONE RESPONSE.
CC POLARIZATION OF YEAST CELLS IN RESPONSE TO MATING PHEROMONES. SOME
CC FAR1 MUTANTS APPEAR TO BE DEFECTIVE IN MATING BECAUSE THEY ARE

CC UNABLE TO LOCATE THE MATING PARTNER.
CC -1- SUBUNIT: ASSOCIATES WITH THE CDC28-CLN COMPLEX.
CC -1- INDUCTION: BY ALPHA-FACTOR IN A CELLS.
CC -1- DOMAIN: THERE IS EVIDENCE TO SUGGEST THAT THE N-TERMINAL MAY BE
CC SUFFICIENT FOR CELL CYCLE ARREST AND THE C-TERMINAL MAY BE
CC NECESSARY FOR SOME STEP IN MATING.
CC -1- PTM: THOUGHT TO BE PHOSPHORYLATED BY MAP KINASE FUS3. THOUGHT TO
CC ENHANCE THE BINDING OF FAR1 TO G1-SPECIFIC CYCLIN-DEPENDENT KINASE
CC (CDK) COMPLEXES.
CC -1- SIMILARITY: TO YEAST STES.
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M60071; AAA34600.1; ALT_INIT.
CC EMBL: 249432; CAA89452.1; -.
CC PIR: S13341; S13341.
CC SGD: S0003693; FAR1.
CC InterPro: IPR001841; znf_ring.
CC SMART: SM00184; RING.
CC PROSITE: PS00518; ZF_RING_1; FALSE_NEG.
CC PROSITE: PS00889; ZF_RING_2; 1.
CC Protein kinase inhibitor; Cell cycle; Cell division; Phosphorylation;
CC Zinc-finger.
KW 2N_FING 202 252
FT MOD_RES 87 87
FT MOD_RES 306 306
FT VARIANT 87 87
FT MUTAGEN 87 87
FT MUTAGEN 306 306
FT CONFLICT 20 20
FT CONFLICT 568 568
SQ SEQUENCE 830 AA: 94572 MW: 834EBB2D7BB964ID CRC64;

Query Match 62.5%; Score 30; DB 1; Length 830;
Best local similarity 71.4%; Pred. No. 1.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLNAVK 9
Db 751 PLNASIK 757

RESULT 13
ACAL_ARATH
ID ACAL_ARATH STANDARD; PRT; 1020 AA.
AC Q37145; Q37146; Q42571; Q42587; Q9SFY1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Calcium-transporting ATPase 1, plasma membrane-type (EC 3.6.3.8)
DE (Ca2+-ATPase, isoform 1) (Plastid envelope ATPase 1).
GN ACAL OR PEA1 OR ATIG27770 OR T22C5.23 OR F28H5.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eudicots II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxId=3702;
RN [1]
RP PRELIMINARY SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=cv. Columbia;
RX MEDLINE=94052104; PubMed=8234257;
RA Huang L., Berkeiman T., Franklin A.E., Hoffman N.E.;
RT "Characterization of a gene encoding a Ca(2+)-ATPase-like protein in
RT the plastid envelope.";

RL [Proc. Natl. Acad. Sci. U.S.A. 90:10066-10070(1993)].
 RN [2].
 RP REVISIONS.
 RA Huang L., Berkelman T., Franklin A.E., Hoffman N.E.;
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9664-9664(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 RA White O., Alonso J., Altati H., Azaio R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Guller J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin X., Liu S.X., Liu Z.A., Lueros J.S., Maltl R., Marshall A.,
 RA Maltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pail G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sancho H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sunk H., Tallon L.J., Tambunga G., Totimni M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant *Arabidopsis*
 RT *thaliana*." ;
 RL Nature 408:816-820(2000).
 CC -I- FUNCTION: THIS MAGNESIUM DEPENDENT ENZYME CATALYZES THE HYDROLYSIS
 CC OF ATP COUPLED WITH THE TRANSLOCATION OF CALCIUM FROM THE CYTOSOL
 CC OUT OF THE CELL OR INTO ORGANELLS.
 CC -I- CATALYTIC ACTIVITY: ATP + H(2)O + Ca(2+)(In) = ADP + phosphate +
 CC Ca(2+)(Out).
 CC -I- ENZYME REGULATION: ACTIVATED BY CALMODULIN (BY SIMILARITY).
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHLOROPLAST
 CC ENVELOPE (INNER MEMBRANE) (PROBABLE).
 CC -I- TISSUE SPECIFICITY: EXPRESSED AT HIGHER LEVELS IN ROOTS THAN IN
 CC LEAVES.
 CC -I- DOMAIN: THE N-TERMINUS CONTAINS AN AUTOINHIBITORY CALMODULIN-
 CC BINDING DOMAIN, WHICH BINDS CALMODULIN IN A CALCIUM-DEPENDENT
 CC FASHION (BY SIMILARITY).
 CC -I- SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY
 CC (EI-E2 ATPASES). SUBFAMILY IIB.
 CC -I- CAUTION: Ref.3 sequence differs from that shown due to erroneous
 CC gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)
 CC or send an email to license@sib-sib.ch.
 CC -----
 DR EMBL, L08468; AAD10211.1; -;
 DR EMBL, L08469; AAD10212.1; -;
 DR EMBL, D13983; BAA03090.1; -;
 DR EMBL, D13984; BAA03091.1; -;
 DR EMBL, X69940; CAA49558.1; -;
 DR EMBL, X69941; CAA49559.1; -;
 DR EMBL, AC012375; AAF24958.1; AUT_SEQ.
 DR EMBL, AC079280; AAG50579.1; -;
 DR HSSP; P04191; IEUL.
 DR InterPro: IPR001757; ATPase_E1-E2.
 DR InterPro: IPR004014; Cation_ATPase.
 DR InterPro: IPR001454; Hlgnaase/hydratase.
 DR Pfam; PF00123; E1-E2_ATPase; 1.
 DR Pfam; PF00689; Cation_ATPase_C; 1.
 DR Pfam; PF00690; Cation_ATPase_N; 1.
 DR Pfam; PF00702; Hydrolase; 1.
 DR PRINTS; PR00119; CARATPASE.
 DR TIGRFAWS; TIGR01116; Ca_ATPase; 1.
 DR PROSITE; PS00154; ATPASE_E1_E2; 1.
 DR

| | |
|------------|--|
| KW | Hydroxylase;Calcium transport; Transmembrane; Phosphorylation; |
| KW | Atp-binding; Metal-binding; Magnesium; Chloroplast; |
| KW | Calmodulin-binding; Multigene family. |
| FT | DOMAIN 1 162 CYTOPLASMIC (POTENTIAL). |
| FT | TRANSSEM 183 POTENTIAL. |
| FT | DOMAIN 184 201 LUMENAL (POTENTIAL). |
| FT | TRANSSEM 202 222 POTENTIAL. |
| FT | DOMAIN 223 350 CYTOPLASMIC (POTENTIAL). |
| FT | TRANSSEM 351 370 POTENTIAL. |
| FT | DOMAIN 371 400 LUMENAL (POTENTIAL). |
| FT | TRANSSEM 401 418 POTENTIAL. |
| FT | DOMAIN 419 813 CYTOPLASMIC (POTENTIAL). |
| FT | TRANSSEM 814 832 POTENTIAL. |
| FT | DOMAIN 833 843 LUMENAL (POTENTIAL). |
| FT | TRANSSEM 844 864 POTENTIAL. |
| FT | DOMAIN 865 884 CYTOPLASMIC (POTENTIAL). |
| FT | TRANSSEM 885 907 POTENTIAL. |
| FT | DOMAIN 908 919 LUMENAL (POTENTIAL). |
| FT | TRANSSEM 920 941 POTENTIAL. |
| FT | DOMAIN 942 959 CYTOPLASMIC (POTENTIAL). |
| FT | TRANSSEM 960 981 POTENTIAL. |
| FT | DOMAIN 982 991 LUMENAL (POTENTIAL). |
| FT | TRANSSEM 992 1013 POTENTIAL. |
| FT | DOMAIN 1014 1020 CYTOPLASMIC (POTENTIAL). |
| FT | DOMAIN 21 32 CALMODULIN-BINDING (BY SIMILARITY). |
| FT | MOD_RES 46 46 PHOSPHORYLATION (BY GDPK) (BY SIMILARITY). |
| FT | MOD_RES 456 456 PHOSPHORYLATION (BY SIMILARITY). |
| FT | METAL 758 758 MAGNESIUM (BY SIMILARITY). |
| FT | METAL 762 762 MAGNESIUM (BY SIMILARITY). |
| FT | CONFLICT 88 88 P -> S (IN REF.1; AAD10211 AND 2; BAA03090/CAA49559). |
| FT | CONFLICT 801 801 I -> T (IN REF. 2; BAA03091/CAA49558 AND 3). |
| SQ | SEQUENCE 1020 AA; 111273 MW; 18A94DE5C1AAFP4F CRC64; |
| QY | Query Match 62.5%; Score 30; DB 1; Length 1020; |
| DB | Best Local Similarity .66.7%; Pred.No. 1.4e+02; |
| DB | Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0 |
| OY | 2 MELNAAVKM 10 : : Db 1006 MPVAAAKM 1014 |
| RESULT 14 | |
| TIE1_MOUSE | STANDARD: PRt: 1134 AA. |
| ID | TIE1_MOUSE |
| AC | 006806; |
| DT | 01-FEB-1995 (Rel. 31, Created) |
| DT | 01-FEB-1996 (Rel. 33, Last sequence update) |
| DT | 15-JUN-2002 (Rel. 41, Last annotation update) |
| DE | Tyrosine-protein kinase receptor Tie-1 precursor (EC 2.7.1.112). |
| GN | TIE1 OR TIE-1. |
| OS | Mus musculus (Mouse). |
| OC | Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; |
| CC | Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. |
| OX | NCBI_TaxID=10090; |
| RN | [1] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=BALB/c; TISSUE=Lung; |
| RX | MEDLINE=9402374; PubMed=8415706; |
| RA | Sato T.N., Qin Y., Kozak C.A., Andus K.L.; |
| RT | "Tie-1 and tie-2 define another class of putative receptor tyrosine kinase genes expressed in early embryonic vascular system."; |
| RL | Proc. Natl. Acad. Sci. U.S.A. 90:9355-9358(1993). |
| RN | [2] |
| RP | SEQUENCE FROM N.A. |
| RC | STRAIN=BALB/c; TISSUE=Bone marrow; |
| RA | Krivsov A.V., Erschler M.A., Visser J.W.M., Belyavsky A.V.; |
| RL | Submitted (JUN-1994) to the EMBL/Genbank/DBJ databases. |
| RN | [3] |
| RP | SEQUENCE FROM N.A. |

RC STRAIN-BALB/C; TISSUE-Liver;
 RX MEDLINE-93371421; PubMed-8395828;
 RA Iwama A., Hamaguchi I., Hashiyama M., Murayama Y., Yasunaga K.,
 RA Suda T.,
 RT "Molecular cloning and characterization of mouse TIE and TEK receptor
 RT tyrosine kinase genes and their expression in hematopoietic stem
 RT cells.";
 RL Biochem. Biophys. Res. Commun. 195:301-309(1993).
 RN [4]
 RP SEQUENCE OF 1-19 FROM N.A.
 RC STRAIN-129/SV; TISSUE-Liver;
 RX MEDLINE-95383653; PubMed-7655012;
 RA Korhonen J., Lahtinen I., Halmekyto M., Alhonen L., Janne J.,
 RA Dumont D., Allitalo K.,
 RT "Endothelial-specific gene expression directed by the tie gene
 RT promoter in vivo.";
 RL Blood 86:1828-1835(1995).
 CC -1- FUNCTION: PROBABLE PROTEIN TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN DEVELOPING VASCULAR
 CC ENDOTHELIAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. TIE
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL: X71425; CAA50556.1; -;
 CC DR EMBL: X80764; CAA56739.1; -;
 CC DR EMBL: X73960; CAA52148.1; -;
 CC DR EMBL: S79346; -; NOT_ANNOTATED_CDS.
 CC HSPSP: P11362.1; FPK.
 CC MGD: MGI:99906; T1e1.
 CC DR InterPro: IPR000561; EGF-like.
 CC DR InterPro: IPR000719; Euk_pkinase.
 CC DR InterPro: IPR003961; FN_III.
 CC DR InterPro: IPR003599; Ig.
 CC DR InterPro: IPR003006; Ig_MHC.
 CC DR InterPro: IPR003600; Ig_Like.
 CC DR InterPro: IPR01245; Tyr_pkinase.
 CC Pfam: PF00008; EGF; 2.
 CC Pfam: PF00041; fn3; 3.
 CC Pfam: PF00047; Ig; 2.
 CC Pfam: PF00069; pkinase; 1.
 CC PRINTS: PR00109; TYR_KINASE.
 CC DR ProDom: PD000001; Euk_pkinase; 1.
 CC DR SMART: SM00181; EGF; 2.
 CC DR SMART: SM00060; FN3; 2.
 CC DR SMART: SM00409; IG; 1.
 CC DR SMART: SM00410; IG_Like; 1.
 CC DR SMART: SM00219; TYR_K; 1.
 CC DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 CC DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 CC DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 CC DR PROSITE: PS00022; EGF_1; 3.
 CC DR PROSITE: PS01186; EGF_2; 3.
 CC Receptor; Tyrosine-protein kinase; Transferase; Signal; ATP-binding;
 CC Repeat; EGF-like domain; Transmembrane; Immunoglobulin domain;
 CC Glycoprotein; Phosphorylation; Multigene family.
 CC SIGNAL 1 22
 CC FT CHAIN 23 1134 TYROSINE-PROTEIN KINASE RECEPTOR TIE-1.
 CC FT DOMAIN 23 755 EXTRACELLULAR (POTENTIAL).
 CC FT TRANSMEM 756 780 POTENTIAL.

FT DOMAIN 781 1134 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 43 103 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 212 254 EGF-LIKE 1.
 FT DOMAIN 256 301 EGF-LIKE 2.
 FT DOMAIN 303 343 EGF-LIKE 3.
 FT DOMAIN 370 424 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 444 538 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 541 635 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 639 740 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 835 1114 PROTEIN KINASE.
 FT NP_BIND 841 849 ATP (BY SIMILARITY).
 FT BINDING 866 866 ATP (BY SIMILARITY).
 FT ACT_SITE 975 975 BY SIMILARITY.
 FT MOD_RES 1003 1003 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
 FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 159 159 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 592 592 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 705 705 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 599 599 R -> L (IN REF. 1).
 SQ SEQUENCE 1134 AA: 124698 MW: 0F212ED6C50ACAE CRC64;
 Query Match 62.5%; Score 30; DB 1; Length 1134;
 Best Local Similarity 55.6%; Pred. No. 1.6e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 2 MFLNAVKM 10
 Db 859 LKMAAIKM 867
 RESULT 15
 T1E1_BOVIN STANDARD; PRT; 1136 AA.
 ID T1E1_BOVIN
 AC 006805;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase receptor Tie-1 precursor (BC 2.7.1.112).
 GN TIE1 OR TIE-1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 ON NCBI_Taxid-9913;
 OX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Endothelial cells;
 RX MEDLINE-94022374; PubMed-8415706;
 RA Sato T.N., Qin Y., Kozak C.A., Andus K.L.;
 RT "Tie-1 and tie-2 define another class of putative receptor tyrosine
 RT kinase genes expressed in early embryonic vascular system.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:9355-9358(1993).
 CC -1- FUNCTION: PROBABLE PROTEIN TYROSINE-KINASE TRANSMEMBRANE RECEPTOR.
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine -> ADP + protein
 CC tyrosine phosphate.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN DEVELOPING VASCULAR
 CC ENDOTHELIAL CELLS.
 CC -1- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. TIE
 CC SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 25, 2003, 08:19:47 ; Search time 22.1212 Seconds
(Without alignments)
93.145 Million cell updates/sec

Title: US-09-646-532B-1
Perfect score: 48
Sequence: 1 SMPLNAVKM 10

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rviro:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
| 1 | 37 | 77.1 | 154 | 16 | Q98MA8 |
| 2 | 36 | 75.0 | 176 | 4 | Q16051 |
| 3 | 36 | 75.0 | 1009 | 4 | Q14527 |
| 4 | 36 | 75.0 | 1009 | 4 | Q96KM9 |
| 5 | 36 | 75.0 | 1009 | 4 | Q14536 |
| 6 | 35 | 72.9 | 836 | 6 | Q95217 |
| 7 | 35 | 72.9 | 1005 | 6 | Q95216 |
| 8 | 34 | 70.8 | 450 | 16 | Q8RDK9 |
| 9 | 34 | 70.8 | 1134 | 16 | Q92RW9 |
| 10 | 33 | 68.8 | 154 | 16 | Q92057 |
| 11 | 33 | 68.8 | 459 | 16 | Q99RU7 |
| 12 | 33 | 68.8 | 487 | 13 | Q90215 |
| 13 | 33 | 68.8 | 639 | 2 | Q87538 |
| 14 | 33 | 68.8 | 1670 | 10 | Q9LE42 |
| 15 | 32 | 66.7 | 154 | 16 | Q8YH89 |
| 16 | 32 | 66.7 | 323 | 10 | Q9XIH8 |

| | | | | | |
|----|----|------|------|----|--------|
| 17 | 32 | 66.7 | 443 | 16 | Q83067 |
| 18 | 32 | 66.7 | 1104 | 5 | Q966J8 |
| 19 | 32 | 66.7 | 1248 | 10 | Q9SNA2 |
| 20 | 31 | 64.6 | 114 | 10 | Q9M083 |
| 21 | 31 | 64.6 | 181 | 16 | Q9JR25 |
| 22 | 31 | 64.6 | 181 | 16 | Q9JR57 |
| 23 | 31 | 64.6 | 279 | 10 | Q80843 |
| 24 | 31 | 64.6 | 301 | 16 | P73681 |
| 25 | 31 | 64.6 | 315 | 16 | Q9K703 |
| 26 | 31 | 64.6 | 348 | 17 | Q9HPV6 |
| 27 | 31 | 64.6 | 364 | 16 | Q9YH80 |
| 28 | 31 | 64.6 | 417 | 2 | Q9F708 |
| 29 | 31 | 64.6 | 421 | 16 | Q9X296 |
| 30 | 31 | 64.6 | 425 | 16 | Q91409 |
| 31 | 31 | 64.6 | 435 | 16 | Q9Z379 |
| 32 | 31 | 64.6 | 435 | 16 | Q25942 |
| 33 | 31 | 64.6 | 451 | 3 | Q8X0M7 |
| 34 | 31 | 64.6 | 454 | 16 | Q8UAS3 |
| 35 | 31 | 64.6 | 503 | 10 | Q9SX12 |
| 36 | 31 | 64.6 | 506 | 11 | Q9CW17 |
| 37 | 31 | 64.6 | 509 | 11 | Q88548 |
| 38 | 31 | 64.6 | 509 | 11 | P97943 |
| 39 | 31 | 64.6 | 509 | 11 | P97943 |
| 40 | 31 | 64.6 | 567 | 3 | Q9U057 |
| 41 | 31 | 64.6 | 650 | 10 | Q49570 |
| 42 | 31 | 64.6 | 666 | 17 | Q27609 |
| 43 | 31 | 64.6 | 707 | 2 | Q9X5R7 |
| 44 | 31 | 64.6 | 716 | 5 | Q9W2D2 |
| 45 | 31 | 64.6 | 726 | 3 | Q96UR7 |

ALIGNMENTS

RESULT 1

| ID | Q98MA8 | PRELIMINARY: | PRT: | 154 AA. |
|----|---|---|------|---------|
| AC | Q98MA8: | 01-OCT-2001 (TRENBLREL. 18, last sequence update) | | |
| DT | 01-OCT-2001 (TRENBLREL. 18, last sequence update) | | | |
| DT | 01-MAR-2002 (TRENBLREL. 20, last annotation update) | | | |
| DE | Hypothetical protein mlr0658. | | | |
| GN | MLR0658. | | | |
| OS | Rhizobium loti (Mesorhizobium loti). | | | |
| OC | Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group; | | | |
| OC | Phyllobacteriaceae; Mesorhizobium. | | | |
| OX | NCBI_TaxID=381; | | | |
| RN | [1] | | | |
| RP | SEQUENCE FROM N.A. | | | |
| RC | STRAIN=MAFE03099; | | | |
| RX | MEDLINE=21082930; PubMed=11214968; | | | |
| RA | Kaneko T., Nakamura Y., Sato S., Asamiizu E., Kato T., Sasamoto S., | | | |
| RA | Watanabe A., Idegawa K., Ishikawa K., Kimura T., | | | |
| RA | Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., | | | |
| RA | Mochizuki Y., Nakayama S., Nakazaki N., Shlimpo S., Sugimoto M., | | | |
| RA | Takeuchi C., Yamada M., Tabata S.; | | | |
| RT | "Complete genome structure of the nitrogen-fixing symbiotic bacterium | | | |
| RT | Mesorhizobium loti." | | | |
| RL | DNA Res. 7:331-338(2000). | | | |
| DR | EMBL; AP002995; BAB48205.1; . | | | |
| DR | InterPro: IPR000543; YJGF-1like. | | | |
| DR | Pfam: PF01042; UPF0076; 1. | | | |
| KW | Hypothetical protein; Complete proteome. | | | |
| SO | SEQUENCE 154 AA; 16078 MW; 3EB53A511967BEO CXC64; | | | |

Query Match 77.1%; Score 37; DB 16; Length 154;
Best Local Similarity 70.0%; Pred. No. 4.3;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SMPLNAVKM 10
DB 138 SLPLNAVEI 147

RESULT 2

016051 PRELIMINARY; PRT; 176 AA.

AC 016051;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE DNA-binding protein/plasminogen activator inhibitor-1 regulator (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 MEDLINE=93342870; PubMed=8342330;
 RA Descheemaeker K.;
 RT "On the regulation of the plasminogen activator inhibitor-1 gene expression";
 RL Koninklijke Acad. Geneeskunde België 55:225-264(1993).
 DR EMBL; S64671; AAB27691.1; -
 KW DNA-binding.
 FT NON_TER
 SQ SEQUENCE 176 AA; 19362 MW; 2D985F5D180AEE21 CRC64;

Query Match

Best Local Similarity 70.0%; Score 36; DB 4; Length 176;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPUNAANK 10
 DB 159 SMPVHAANK 168

014527 PRELIMINARY; PRT; 1009 AA.

ID 014527;
 AC 014527;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE ATPase.
 GN HPI16.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 MEDLINE=95181452; PubMed=7876228;
 RA Sheridan P.L., Schorpp M., Voz M.L., Jones K.A.;
 RT "Cloning of an SNF2/SWI2-related protein that binds specifically to the SPH motifs of the SV40 enhancer and to the HIV-1 promoter";
 RL J Biol Chem. 270:4575-4587(1995)
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 DE EMBL; U34673; AAA67436.1; -
 DR InterPro: IPR001410; DEAD.
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR InterPro: IPR001841; znf_fing.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00176; SNF2_N; 2.
 DR SMART: SM00487; zf-C3HC4; 1.
 DR SMART: SM00490; HelicC; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00518; zf_RING_1; 1.
 KW ATP-binding; DNA-binding; Helicase; Zinc-finger.
 SQ SEQUENCE 1009 AA; 113942 MW; 0AB80F4B6AA72645 CRC64;

Query Match 75.0%; Score 36; DB 4; Length 1009;
 Best Local Similarity 70.0%; Pred. No. 45;

Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPUNAANK 10
 DB 196 SMPVHAANK 205

RESULT 4

096KM9 PRELIMINARY; PRT; 1009 AA.

AC 096KM9;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE SWI/SNF related, matrix associated, actin dependent regulator of DE chromatin, subfamily a, member 3.
 GN SMARCA3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RA Ribaucour F., Wiedig M., Benoitmane A.M., Coppée F., Belayew A.;
 RT "Characterization of the human SMARCA3/HLTF gene";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
 DR EMBL; A418064; CAD10805.1; -
 DR InterPro: IPR001650; Helicase_C.
 DR InterPro: IPR000330; SNF2_N.
 DR InterPro: IPR001841; znf_fing.
 DR Pfam: PF00271; helicase_C; 1.
 DR Pfam: PF00176; SNF2_N; 1.
 DR SMART: SM00487; zf-C3HC4; 1.
 DR PROSITE: PS00518; zf_RING_1; UNKNOWN_1.
 KW ATP-binding; Helicase; Zinc-finger.
 SQ SEQUENCE 1009 AA; 113898 MW; 311E4A8A8F83782A CRC64;

Query Match

Best Local Similarity 70.0%; Score 36; DB 4; Length 1009;
 Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPUNAANK 10
 DB 196 SMPVHAANK 205

RESULT 5

014536 PRELIMINARY; PRT; 1009 AA.

ID 014536;
 AC 014536; Q92652;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Helicase-like transcription factor.
 GN HLTF-1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 MEDLINE=96268471; PubMed=8672239;
 RA Ding H., Descheemaeker K., Marynen P., Nelles L., Carvalho T.,
 RA Carmo-Fonseca M., Collen D., Belayew A.;
 RT "Characterization of a helicase-like transcription factor involved in RT the expression of the human plasminogen activator inhibitor-1 gene";
 RL DNA Cell Biol. 15:429-442(1996).
 [2]
 SQ SEQUENCE OF 123-1009 FROM N.A.
 MEDLINE=95181452; PubMed=7876228;
 RA Sheridan P.L., Schorpp M., Voz M.L., Jones K.A.;
 RT "Cloning of an SNF2/SWI2-related protein that binds specifically to

RT the SPH motifs of the SV40 enhancer and to the HIV-1 promoter.";
RL J. Biol. Chem. 270:4575-4587(1995).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: Z46606; CA86571.1; -;
DR EMBL: Z46606; CA86572.1; -;
DR TRANSFAC: T04146; -;
DR TRANSFAC: T04147; -;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 2.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR ATP-binding; Helicase; Zinc-finger.
SQ SEQUENCE 1009 AA; 113914 MW; B4ACD86A6844CC5F CRC64;

Query Match 75.0%; Score 36; DB 4; Length 1009;
Best Local Similarity 70.0%; Pred. No. 45;
Matches 7; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPLNAVKM 10
||||:||||
DB 196 SMPVHAQVM 205

RESULT 6
O95217 PRELIMINARY; PRT; 836 AA.
AC O95217;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RUSH-1beta.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE;
RX MEDLINE-97082217; PubMed-8923460;
RA Hayward-Lester A., Hewatson A., Beale E.G., Oefner P.J., Doris P.A.,
RA Chilton B.S.;
RT "Cloning, characterization, and steroid-dependent posttranscriptional
RT processing of RUSH-1 alpha and beta, two uteroglobin promoter-binding
RT proteins.";
RL Mol. Endocrinol. 10:1335-1349(1996).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: U66565; AAC48693.1; -;
DR HSSP: P15919; IRMD.
DR TRANSFAC: T04163; -;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR Zinc-finger.
SQ SEQUENCE 836 AA; 94841 MW; B0ED6969548035DD CRC64;

Query Match 72.9%; Score 35; DB 6; Length 836;
Best Local Similarity 60.0%; Pred. No. 62;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
OY 1 SMPLNAVKM 10
||||:||||

DB 196 SMPVHAQVM 205

RESULT 7
O95216 PRELIMINARY; PRT; 1005 AA.
AC O95216;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE RUSH-1alpha.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW ZEALAND WHITE;
RX MEDLINE-97082217; PubMed-8923460;
RA Hayward-Lester A., Hewatson A., Beale E.G., Oefner P.J., Doris P.A.,
RA Chilton B.S.;
RT "Cloning, characterization, and steroid-dependent posttranscriptional
RT processing of RUSH-1 alpha and beta, two uteroglobin promoter-binding
RT proteins.";
RL Mol. Endocrinol. 10:1335-1349(1996).
CC -1- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL: U66564; AAC18656.1; -;
DR HSSP: P15919; IRMD.
DR TRANSFAC: T04162; -;
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; Znf_fing.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00176; SNF2_N; 1.
DR Pfam: PF00097; zf-C3HC4; 1.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00184; RING; 1.
DR PROSITE: PS00518; ZF_RING_1; 1.
DR ATP-binding; Helicase; Zinc-finger.
SQ SEQUENCE 1005 AA; 113582 MW; C741E7117D6BD807 CRC64;

Query Match 72.9%; Score 35; DB 6; Length 1005;
Best Local Similarity 60.0%; Pred. No. 74;
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPLNAVKM 10
||||:||||
DB 196 SMPVHAQVM 205

RESULT 8
O8BDK9 PRELIMINARY; PRT; 450 AA.
AC O8BDK9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE UDP-N-acetylmuramyl tripeptide synthase.
GN MURE OR TRE0008.
OS Thermoaerobacter tengcongensis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridia;
OC Thermoaerobacteriales; Thermoaerobacteriaceae; Thermoaerobacter.
OX NCBI_TaxID=119072;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MB4T / JCM11007;
RX MEDLINE-21992816; PubMed-11997336;
RA Bao Y., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Q., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";

```

RL Genome Res. 12:689-700(2002).
DR EMBL; AF012974; AAM23325.1; -.
KM Complete proteome.
SQ SEQUENCE 450 AA; 50088 MW; 9DD5C0F7C129A8C1 CRC64;

Query Match
Best Local Similarity 70.8%; Score 34; DB 16; Length 450;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVK 9
   |||||
Db 409 MPLNAVK 416

RESULT 9
O92RM9 PRELIMINARY; PRT; 1134 AA.
AC O92RM9;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative sensor histidine kinase protein (EC 2.7.).
GN R00721 OR SMC00776.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ranspenger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591784; CAC45293.1; -.
DR InterPro; IPR003594; ATPbind_ATPase.
DR InterPro; IPR004359; HIS_KIN_sig.
DR InterPro; IPR000014; PAS_domain.
DR Pfam; PF02518; HATPase_C; 1.
DR Pfam; PF00989; PAS; 2.
DR Pfam; PF00512; Signal; 1.
DR TIGRFAMS; TIGR00229; sensory_box; 2.
KM Transferase; Complete proteome.
SQ SEQUENCE 1134 AA; 120063 MW; 16ACFF65B2E8240 CRC64;

Query Match
Best Local Similarity 66.7%; Score 34; DB 16; Length 1134;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 MPLNAVK 10
   |||||
Db 98 LPLNAVEL 106

RESULT 10
O92057 PRELIMINARY; PRT; 154 AA.
AC O92057;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Hypothetical protein R01493.
GN R01493 OR SMC02103.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Bolstad P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kiss E., Lelaure V., Masuy D.,
RA Pohl T., Portetelle D., Puehler A., Purnelle B., Ranspenger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591787; CAC46072.1; -.
DR InterPro; IPR000543; YJ9F-like.
DR Pfam; PF01042; UPE0076; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 154 AA; 15940 MW; 13C669F0F462BEF7 CRC64;

Query Match
Best Local Similarity 68.8%; Score 33; DB 16; Length 154;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAVK 10
   |||||
Db 138 SMPLNAVK 147

RESULT 11
O99RU7 PRELIMINARY; PRT; 459 AA.
AC O99RU7;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein SAV2326.
GN SAV2326 OR SA2117.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.aureus (strain Mu50), and S.aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Ihan J.-O., Ito T.,
RA Kanamori M., Matsunaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-T., Kaito C.,
RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003365; BAB58468.1; -.
DR EMBL; AP003136; BAB43418.1; -.
DR InterPro; IPR004770; Antiport_nhaC.
DR Pfam; PF03553; Na_H_antiporter; 1.
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 459 AA; 48951 MW; 7342E143DDDEE31D CRC64;

Query Match
Best Local Similarity 75.0%; Score 33; DB 16; Length 459;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SMPLNAVK 8
   |||||
Db 56 SMPLNAVK 63

RESULT 12
O90Z15 PRELIMINARY; PRT; 487 AA.
ID O90Z15

```

AC Q90215;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
 DE Polycumblike 2.
 GN XPCL2.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 CC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8353;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21359116; PubMed=11466526;
 RA Kitaguchi T., Nakata K., Nagai T., Aruga J., Mikoshiba K.,
 RT "Xenopus Polycumblike 2 (XPCL2) controls anterior to posterior
 RL patterning of the neural tissue."
 RL Dev. Genes Evol. 211:309-314(2001).
 DR EMBL; AB050540; BAB43943.1;
 DR InterPro: IPR001965; znf_PHD.
 DR Pfam: PF00628; PHD; 2.
 SQ SEQUENCE 487 AA; 55523 MW; C1FFB52F66DF50C CRC64;

Query Match 68.8%; Score 33; DB 13; Length 487;
 Best Local Similarity 77.8%; Pred. No. 99;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 2 MPLNAVKM 10
 Db 353 VPPNAVKM 361

RESULT 13
 087538 PRELIMINARY; PRT; 639 AA.
 AC 087538;
 DT 01-NOV-1998 (TReMBLrel. 08, Created)
 DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Decca-heme c-type cytochrome.
 GN MPRF.
 OS Shewanella putrefaciens (Pseudomonas putrefaciens).
 CC Bacteria; Proteobacteria; gamma subdivision; Alteromonadaceae;
 CC Shewanella.
 OX NCBI_TaxID=24;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MR-1;
 RA Beliaev A.S.; Saffarini D.A.;
 RT "Isolation of a multi-cytochrome gene cluster mtrDEF involved in metal
 RT reduction in Shewanella putrefaciens."
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF083240; AAD05536.1;
 DR InterPro: IPR001064; Crystallin.
 DR PROSITE: PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
 SQ SEQUENCE 639 AA; 67460 MW; 8E174D9402D39FAD CRC64;

Query Match 68.8%; Score 33; DB 2; Length 639;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 SMPLNAVK 9
 Db 561 ALPLNAVK 569

RESULT 14
 09LE42 PRELIMINARY; PRT; 1670 AA.
 AC 09LE42;
 DT 01-OCT-2000 (TReMBLrel. 15, Created)
 DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)

DE CREB-binding protein-like.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RA Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-COLUMBIA;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Nakamura Y.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty pl and YAC
 RT clones."
 RL DNA Res. 7:131-135(2000).
 DR EMBL; AB026645; BAB02507.1;
 DR InterPro: IPR000345; CYC_heme_bind.
 DR InterPro: IPR000197; TAZ_finger.
 DR InterPro: IPR000822; znf_C2H2.
 DR InterPro: IPR001965; znf_PHD.
 DR InterPro: IPR000433; znf_zz.
 DR Pfam: PF00628; PHD; 1.
 DR Pfam: PF02135; zf-TAZ; 2.
 DR SMART; SM00249; znf_zz; 2.
 DR SMART; SM00291; znf_zz; 2.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR PROSITE: PS01357; zf_zz_1; UNKNOWN_2.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
 SQ SEQUENCE 1670 AA; 186187 MW; 2120E9184B4B98BD CRC64;

Query Match 68.8%; Score 33; DB 10; Length 1670;
 Best Local Similarity 70.0%; Pred. No. 3.3e+02;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 SMPLNAVKM 10
 Db 269 SIPLNAGVM 278

RESULT 15
 08YHH9 PRELIMINARY; PRT; 154 AA.
 AC 08YHH9;
 DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Translation Initiation Inhibitor.
 GN BME10820.
 OS Brucella melitensis.
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA Delvecchio V.G., Kapral V., Redkar R.J., Patra G., Mufer C., Los T.,
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Telleson J.-J.,
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL; AE009523; AAL52001.1;
 DR InterPro: IPR000543; Y39F-like.
 DR Pfam: PF01042; DPF0076; 1.
 KW Complete proteome.

SQ SEQUENCE 154 AA; 15809 MM; EB4DAC4B319FE239 CRC64;
 Query Match 66.7%; Score 32; DB 16; Length 154;
 Best Local Similarity 60.0%; Pred. No. 53;
 Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SMPLNAAVKM 10
 1:1111111:
 Db 138 SLPLNAPVEV 147

Search completed: March 25, 2003, 08:21:33
 Job time : 26.1212 secs